



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

**Συσταδοποίηση Κινησιογραμμάτων Ελληνικού  
Παραδοσιακού Χορού με χρήση τεχνικών Βαθιάς Μάθησης**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

**ΑΓΓΕΛΙΚΗΣ – ΕΛΕΝΗΣ ΔΟΥΚΙΔΗ**

**Επιβλέπων:** Ανδρέας-Γεώργιος Σταφυλοπάτης

Καθηγητής Ε.Μ.Π.

**Συνεπιβλέπουσα:** Κατερίνα Ελ Ράχεμπ

Επίκουρη Καθηγήτρια ΠΑ.ΠΕΛ.

---

Αθήνα, Νοέμβριος 2023





ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

ΕΡΓΑΣΤΗΡΙΟ ΣΥΣΤΗΜΑΤΩΝ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΚΑΙ ΜΑΘΗΣΗΣ

## Συσταδοποίηση Κινησιογραμμάτων Ελληνικού Παραδοσιακού Χορού με χρήση τεχνικών Βαθιάς Μάθησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

**ΑΓΓΕΛΙΚΗΣ – ΕΛΕΝΗΣ ΔΟΥΚΙΑΔΗ**

**Επιβλέπων:** Ανδρέας-Γεώργιος Σταφυλοπάτης

Καθηγητής Ε.Μ.Π.

**Συνεπιβλέπουσα:** Κατερίνα Ελ Ράχεμπ

Επίκουρη Καθηγήτρια ΠΑ.ΠΕΛ.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 10<sup>η</sup> Νοεμβρίου 2023.

.....  
Ανδρέας-Γεώργιος Σταφυλοπάτης

Καθηγητής Ε.Μ.Π.

.....  
Γεώργιος Στάμου

Καθηγητής Ε.Μ.Π.

.....  
Κατερίνα Ελ Ράχεμπ

Επίκουρη Καθηγήτρια ΠΑ.ΠΕΛ.

Αθήνα, Νοέμβριος 2023



.....

## **ΑΓΓΕΛΙΚΗ - ΕΛΕΝΗ ΔΟΥΚΙΔΗ**

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

Copyright © - All rights reserved Αγγελική – Ελένη Δουκίδη, 2023.

Με επιφύλαξη παντός δικαιώματος.

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

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



# Περίληψη

---

Ο Χορός είναι μια μορφή τέχνης, μια αντανάκλαση συναισθημάτων και ιδεών στην κίνηση του σώματος. Αποτελούσε ανέκαθεν ένα μέσο επικοινωνίας και ανταλλαγής, μεταβίβασης γνώσεων και εθιμοτυπικών παραδόσεων, αλλά και συνάμα ψυχαγωγίας των ανθρώπων όλων των κοινωνιών. Ανάμεσα στα πολυάριθμα είδη χορών που υπάρχουν, ωστόσο, ξεχωρίζει ο Ελληνικός Παραδοσιακός Χορός, μια πολιτισμική κληρονομιά ανεκτίμητης αξίας που καθρεφτίζει την ελληνική κουλτούρα και διασφαλίζει τη διατήρηση της διαφορετικότητας σε μια περίοδο πολιτισμικής ομογενοποίησης. Για τη μετάδοσή του, όμως, από γενιά σε γενιά, αναγκαία κρίνεται η ύπαρξη γραπτών τεκμηρίων, με άλλα λόγια, η καταγραφή του. Ανά τα χρόνια, έχουν δημιουργηθεί διαφορετικά συστήματα σημειογραφίας της χορευτικής κίνησης, σε μια προσπάθεια όσο το δυνατόν ακριβέστερης αποτύπωσής της στο χαρτί. Αυτές οι αλληλουχίες κινήσεων που πλέον, στην πλειοψηφία τους, συντάσσονται και διατηρούνται σε ψηφιακή μορφή, ονομάζονται κινησιογράμματα.

Παρά τις ραγδαίες εξελίξεις στον τομέα της τεχνολογίας, τα τελευταία μόνο χρόνια η επιστημονική κοινότητα αναγνωρίζει τη σπουδαιότητα και στρέφει το ενδιαφέρον της προς τη μελέτη της καταγραφής του χορού. Πράγματι, αποτελεί ένα πεδίο που βρίσκεται σε φάση εξέλιξης, με εκτεταμένες μελέτες και νέα, πρωτοποριακά πειράματα να λαμβάνουν χώρα σε παγκόσμιο επίπεδο. Με τις σύγχρονες τεχνικές Βαθιάς Μάθησης, δε, να ξεπερνούν κάθε προσδοκία και αξιοποιώντας τις απόθμενες δυνατότητες που μας προσφέρουν, φαίνεται να ανοίγει ο δρόμος προς περαιτέρω εμβάθυνση, ανάλυση και επεξεργασία των κινησιογραμμάτων, τα οποία, βέβαια, αποτελούν ακόμη μεγάλη πρόκληση για τους ειδικούς λόγω της πολυπλοκότητάς τους.

Στην παρούσα διπλωματική εργασία, σε μια προσπάθεια σύνδεσης των επιστημονικών πεδίων της Βαθιάς Μάθησης με τη Σημειογραφία του Χορού, μελετάται ο διαχωρισμός σε συστάδες κινησιογραμμάτων που αναπαριστούν παραδοσιακούς χορούς της Ελλάδας, ανάλογα με την κατηγορία του χορού στην οποία ανήκουν. Αρχικά, γίνεται κατάλληλη προεπεξεργασία των δεδομένων, έτσι ώστε αυτά να χρησιμοποιηθούν σαν είσοδος σε οχτώ διαφορετικές αρχιτεκτονικές Συνελκτικών Νευρωνικών Δικτύων, οι οποίες λειτουργούν ως εξαγωγείς των χρήσιμων χαρακτηριστικών τους. Ακολουθούν τεχνικές μείωσης της διαστατικότητάς τους, ώστε τα στοιχεία αυτά να τροφοδοτηθούν σε αλγόριθμο Συσταδοποίησης, ο οποίος αποφέρει και τα τελικά αποτελέσματα. Η έρευνα ολοκληρώνεται με επιλογή των βέλτιστων αρχιτεκτονικών βάσει της αξιολόγησης των πειραμάτων.

## Λέξεις Κλειδιά

Σημειογραφία Χορευτικής Κίνησης, Κινησιόγραμμα, Παρτιτούρα Χορού, Ελληνικός Παραδοσιακός Χορός, Συσταδοποίηση, Βαθιά Μάθηση, μη Επιβλεπόμενη Μηχανική Μάθηση, Συνελκτικά Νευρωνικά Δίκτυα, Μεταφορά Μάθησης





# Abstract

---

Dance is a form of art, a reflection of emotions and ideas in the movement of the body. It has been a means of communication and knowledge exchange for centuries, preserving cultural heritage and providing entertainment to people across different societies. Among the many dance genres, Greek Folk Dance stands out. It is an invaluable cultural heritage that reflects Greek culture and ensures the preservation of diversity during a period of cultural homogenization. However, the transmission of Greek Folk Dance from generation to generation requires written documentation, a form of dance notation. Over the years, various systems of dance notation have been developed in an attempt to accurately capture dance movements on paper. Nowadays, most of these movement sequences have been organized and preserved in digital formats, commonly referred to as kinetograms.

Despite rapid advancements in technology, only in recent years the scientific community has recognized the importance of the dance notation as research field. It is indeed a field that is evolving, with extensive research and innovative experiments taking place worldwide. With modern Deep Learning techniques exceeding all expectations and utilizing their vast capabilities, it seems that a path towards further in-depth analysis and processing of kinetograms is opening up. However, these kinetograms still present a significant challenge due to their complexity.

In this diploma thesis, an attempt is made, so as to bridge the scientific fields of Deep Learning with Dance Notation. The research focuses on clustering kinetograms representing traditional Greek dances based on their dance category. Initially, the data is preprocessed so as to be used as input for eight different Convolutional Neural Network architectures, which act as feature extractors. Dimensionality reduction techniques follow to prepare the extracted features for clustering. The clustering algorithm is then applied, yielding the final results. The research concludes with the selection of the best architectures based on experimental evaluations.

## Keywords

Dance Notation, Kinetogram, Dance Score, Greek Folk Dance, Clustering, Deep Learning, Unsupervised Machine Learning, Convolutional Neural Networks, Transfer Learning



## Ευχαριστίες

---

Με τη διπλωματική αυτή εργασία ολοκληρώνεται ο προπτυχιακός κύκλος σπουδών μου στη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου. Έτσι, θα ήθελα να εκφράσω τις ειλικρινείς μου ευχαριστίες σε όλους όσους συνέβαλαν σε αυτή την προσπάθεια, παρά τις δυσκολίες της, και ακόμη σε όσους με στήριξαν έμπρακτα και στάθηκαν στο πλευρό μου μέχρι να πετύχω το στόχο μου.

Αρχικά, θα ήθελα να πω ένα μεγάλο ευχαριστώ στον επιβλέποντα καθηγητή κ. Ανδρέα – Γεώργιο Σταφυλοπάτη, ο οποίος με εμπιστεύτηκε από την πρώτη στιγμή, δίνοντάς μου την ευκαιρία να εκπονήσω μια διεπιστημονική εργασία, συμπεριλαμβάνοντας σε αυτή ένα σημαντικό κομμάτι της ζωής μου, το Χορό. Ακόμη, ευχαριστώ ιδιαίτερα τη συνεπιβλέπουσα Δρ. Κατερίνα Ελ Ράχεμπ, επίκουρη καθηγήτρια στο Τμήμα Παραστατικών και Ψηφιακών Τεχνών στη Σχολή Καλών Τεχνών του Πανεπιστημίου Πελοποννήσου, χωρίς τη συμβολή της οποίας η διπλωματική μου εργασία θα ήταν ανέφικτο να πραγματοποιηθεί. Οι πολύτιμες συμβουλές, η καθοδήγηση και οι επισημάνσεις τους ήταν καθοριστικές στη διάρκεια αυτού του ερευνητικού ταξιδιού.

Θα ήθελα, επίσης, να ευχαριστήσω την καθηγήτρια χορού Αθηνά Ανδρουλιδάκη που αποτέλεσε πηγή έμπνευσης αυτής της εργασίας, στην προσπάθειά της να μεταδώσει τις γνώσεις της σχετικά με τη Σημειογραφία του Χορού, καθώς και για την ανιδιοτελή παραχώρηση από πλευράς της χρήσιμων συγγραμμάτων και σημειώσεων που εμπλούτισαν την έρευνά μου.

Φυσικά, δε θα μπορούσα να μη συμπεριλάβω τους γονείς και την αδερφή μου που ήταν πάντοτε στο πλάι μου, όπως και τον Γιώργο που με την υποστήριξή του, μου έδινε δύναμη να συνεχίζω την προσπάθειά μου. Τέλος, θα ήθελα να εκφράσω τη βαθιά μου ευγνωμοσύνη προς τον Δημήτρη, συμφοιτητή και συνοδοιπόρο σε αυτή τη σχολή, για την αμέριστη βοήθεια που μου παρείχε από την αρχή ως το τέλος αυτής της εργασίας. Σας ευχαριστώ για τη συνεχή παρουσία και την αγκαλιά που μου προσφέρατε, στο δρόμο που επέλεξα να διανύσω.

Αθήνα, Νοέμβριος 2023

*Αγγελική - Ελένη Δουκίδη*



# Περιεχόμενα

---

<b>Περίληψη</b> .....	<b>1</b>
<b>Abstract</b> .....	<b>3</b>
<b>Ευχαριστίες</b> .....	<b>5</b>
<b>1 Εισαγωγή</b> .....	<b>17</b>
1.1 Αντικείμενο της Διπλωματικής.....	19
1.2 Κίνητρο της Διπλωματικής.....	20
1.3 Παρεμφερείς Μελέτες.....	21
1.4 Οργάνωση της Διπλωματικής.....	23
<b>I Θεωρητικό Μέρος</b> .....	<b>25</b>
<b>2 Η τέχνη του Χορού</b> .....	<b>27</b>
2.1 Ιστορική αναδρομή.....	27
2.2 Ελληνικός Παραδοσιακός Χορός.....	29
2.3 Σημειογραφία του Χορού.....	31
2.3.1 Labanotation.....	33
2.4 Χορός και Τεχνολογία.....	36
<b>3 Θεωρητικό Υπόβαθρο</b> .....	<b>41</b>
3.1 Τεχνητή Νοημοσύνη.....	41
3.2 Μηχανική Μάθηση.....	42
3.2.1 Επιβλεπόμενη Μάθηση.....	43
3.2.2 Μη Επιβλεπόμενη Μάθηση.....	44
3.2.3 Ενισχυτική Μάθηση.....	46
3.3 Βαθιά Μηχανική Μάθηση.....	46
3.4 Μείωση Διαστατικότητας.....	48
3.4.1 Ανάλυση Κύριων Συνιστωσών.....	49
3.5 Συσταδοποίηση.....	51
3.5.1 Αλγόριθμοι Συσταδοποίησης.....	53
3.5.1.1 Αλγόριθμος k-Means.....	57
3.5.2 Εγκυρότητα συσταδοποίησης.....	58

3.6 Νευρωνικά Δίκτυα.....	59
3.6.1 Τεχνητά Νευρωνικά Δίκτυα .....	61
3.6.2 Perceptron.....	68
3.6.3 Πολυεπίπεδα Νευρωνικά Δίκτυα .....	69
3.6.4 Βαθιά Νευρωνικά Δίκτυα.....	70
3.7 Συνελκτικά Νευρωνικά Δίκτυα.....	72
3.7.1 Επίπεδα επεξεργασίας .....	74
3.8 Μεταφορά Μάθησης .....	79
<b>II Πρακτικό Μέρος.....</b>	<b>83</b>
<b>4 Ανάλυση και Σχεδίαση.....</b>	<b>85</b>
4.1 Δεδομένα: Ανάλυση και Προεπεξεργασία .....	85
4.1.1 Σύνολο Δεδομένων.....	85
4.1.2 Προεπεξεργασία Δεδομένων .....	89
4.1.2.1 Αλλαγή μεγέθους εικόνων.....	89
4.1.2.2 Εξαγωγή χαρακτηριστικών.....	90
4.1.2.3 Ανάλυση Κυρίων Συνιστωσών.....	91
4.2 Σχεδιασμός και Περιγραφή υλοποίησης .....	91
4.2.1 Εργαλεία.....	92
4.2.2 Αρχιτεκτονικές Νευρωνικών Δικτύων .....	93
4.2.2.1 CNN δίκτυα from scratch.....	93
4.2.2.2 VGG .....	96
4.2.2.3 ResNet50 .....	97
4.2.2.4 Inception.....	99
4.2.2.5 Xception .....	101
<b>5 Αξιολόγηση αποτελεσμάτων.....</b>	<b>105</b>
5.1 Μετρικές Αξιολόγησης Συσταδοποίησης.....	106
5.1.1 NMI score.....	106
5.1.2 Rand index score .....	108
5.1.3 Davies-Bouldin Index.....	108
5.1.4 Purity score.....	109
5.2 Αποτελέσματα μη προ-εκπαιδευμένων μοντέλων.....	110
5.2.1 Μοντέλο Πρώτο .....	110
5.2.2 Μοντέλο Δεύτερο .....	113
5.2.3 Μοντέλο Τρίτο .....	115
5.3 Αποτελέσματα προ-εκπαιδευμένων μοντέλων.....	117
5.3.1 VGG16 .....	117

5.3.2 VGG19 .....	119
5.3.3 ResNet50 .....	121
5.3.4 InceptionV3 .....	123
5.3.5 Xception .....	125
5.4 Συγκεντρωτικά αποτελέσματα .....	127
5.5 Μεμονωμένες περιπτώσεις .....	132
5.5.1 Κινησιογράμματα από το ίδιο γένος Χορού.....	132
5.5.1.1 Ο Γιάννης μας αρρώστησε .....	132
5.5.1.2 Γκίκνα.....	134
5.5.2 Κινησιογράμματα από διαφορετικά γένη Χορών.....	134
5.5.2.1 Ομοιότητα ως προς την ακολουθία κινήσεων .....	134
5.5.2.2 Ομοιότητα ως προς τα επαναλαμβανόμενα μοτίβα.....	135
5.5.2.3 Ομοιότητα ως προς το μήκος κινησιογράμματος.....	136
<b>III Επίλογος .....</b>	<b>139</b>
<b>6 Συμπεράσματα και Μελλοντικές Επεκτάσεις .....</b>	<b>141</b>
6.1 Σύνοψη .....	141
6.2 Συμπεράσματα.....	142
6.3 Μελλοντικές επεκτάσεις.....	146
<b>Βιβλιογραφία .....</b>	<b>151</b>





## Κατάλογος Σχημάτων

---

Σχήμα 2.1: Στιγμιότυπο από την παράσταση Grand Finale - Hofesh Shechter Company (2018). (Πηγή [36]) .....	28
Σχήμα 2.2: Συγκαθιστός ή Μαντηλάτος – Παραδοσιακός Χορός Θράκης. (Πηγή [40]).....	30
Σχήμα 2.3: Παράδειγμα καταγραφής χορού με το σύστημα του Feuillet. (Πηγή [45]) .....	32
Σχήμα 2.4: Ονοματοθεσία στηλών – με έντονο χρώμα απεικονίζονται οι τρεις βασικές γραμμές του τρίγραμμου στύλου, ενώ οι υπόλοιπες είναι νοητές. (Πηγή [46]).....	33
Σχήμα 2.5: Σύμβολα που καθορίζουν την κατεύθυνση και το επίπεδο της κίνησης. (Πηγή [47]).....	34
Σχήμα 2.6: Αρχές Labanotation: α) διαδοχικές κινήσεις, β) ταυτόχρονες κινήσεις. (Πηγή [44]) .....	35
Σχήμα 2.7: Κινησιόγραμμα του χορού Ζωναράδικος. ....	36
Σχήμα 2.8: Στιγμιότυπο από την παράσταση Homemade της Trisha Brown (1966). (Πηγή [51]).....	37
Σχήμα 2.9: Χορεύτριες και εικονικές φιγούρες στην παράσταση Biped του Merce Cunningham (1999). (Πηγή [55]).....	38
Σχήμα 3.1: Διάγραμμα συσχέτισης AI – ML – DL.....	47
Σχήμα 3.2: Συγκριτική παρουσίαση της απόδοσης δικτύων Μηχανικής και Βαθιάς Μάθησης. (Πηγή [69]).....	48
Σχήμα 3.3: Ποσοστό περιεχόμενης πληροφορίας ανά κύρια συνιστώσα. (Πηγή [74]) .....	50
Σχήμα 3.4: Intra-cluster και inter-cluster similarity. (Πηγή [76]) .....	51
Σχήμα 3.5: Παράδειγμα συσταδοποίησης με χρήση διαχωριστικών αλγορίθμων. (Πηγή [81]) .....	53
Σχήμα 3.6: Διαδικασία Ιεραρχικής Συσταδοποίησης με συσσωρευτικούς (πάνω) και διαιρετικούς (κάτω) αλγορίθμους. (Πηγή [82]) .....	54
Σχήμα 3.7: Παράδειγμα συσταδοποίησης με βάση την πυκνότητα. (Πηγή [81]).....	55
Σχήμα 3.8: Παράδειγμα συσταδοποίησης με βάση τα πλέγματα. (Πηγή [83]).....	55
Σχήμα 3.9: Παράδειγμα συσταδοποίησης με βάση μοντέλα. (Πηγή [86]) .....	56
Σχήμα 3.10: Εκτέλεση του αλγορίθμου k-means με k=3. (Πηγή [87]).....	57
Σχήμα 3.11: Δομή ενός βιολογικού νευρώνα. (Πηγή [61]).....	60
Σχήμα 3.12: Συναρτήσεις ενεργοποίησης. ....	62

Σχήμα 3.13: Δομή ενός τεχνητού νευρώνα. (Πηγή [95]) .....	63
Σχήμα 3.14: Πλήρως συνδεδεμένα (αριστερά) και μερικώς συνδεδεμένα (δεξιά) δίκτυα. (Πηγή [99]).....	66
Σχήμα 3.15: Δίκτυα πρόσθιας τροφοδότησης (αριστερά) και αναδρομικά δίκτυα (δεξιά). (Πηγή [100]).....	67
Σχήμα 3.16: Το Perceptron. (Πηγή [101]).....	69
Σχήμα 3.17: Απλό vs Βαθύ Νευρωνικό Δίκτυο. (Πηγή [102]) .....	70
Σχήμα 3.18: Συγκριτική παρουσίαση Νευρωνικών Δικτύων. (Πηγή [103]).....	71
Σχήμα 3.19: Συνελικτικό Νευρωνικό Δίκτυο. (Πηγή [105]).....	72
Σχήμα 3.20: Παράδειγμα μίας RGB εικόνας διαστάσεων (4×4×3). (Πηγή [113]) .....	74
Σχήμα 3.21: Συνελικτικό Επίπεδο. (Πηγή [114]).....	76
Σχήμα 3.22: Οι χάρτες χαρακτηριστικών στοιβάζονται και δημιουργούν έναν πίνακα χαρακτηριστικών. (Πηγή [115]).....	76
Σχήμα 3.23: Συγκεντρωτικό επίπεδο με χρήση max pooling. (Πηγή [116]).....	78
Σχήμα 3.24: Σχηματική αναπαράσταση των επιπέδων επεξεργασίας ενός CNN. (Πηγή [117]) .....	79
Σχήμα 3.25: Συγκριτική παρουσίαση της απόδοσης ενός δικτύου που χρησιμοποιεί μεταφορά μάθησης, με δίκτυο που δεν χρησιμοποιεί. (Πηγή [119]) .....	81
Σχήμα 3.26: Παραδοσιακή Μηχανική Μάθηση (αριστερά) και Μεταφορά Μάθησης (δεξιά). (Πηγή [118]).....	82
Σχήμα 4.1: Πλήθος κινησιογραμμάτων σε κάθε γένος χορού.....	87
Σχήμα 4.2: Πρώτο CNN μοντέλο from scratch.....	94
Σχήμα 4.3: Δεύτερο CNN μοντέλο from scratch. ....	95
Σχήμα 4.4: Τρίτο CNN μοντέλο from scratch.....	96
Σχήμα 4.5: Αρχιτεκτονική μοντέλου VGG-16. (Πηγή [128]).....	97
Σχήμα 4.6: Συγκριτική παρουσίαση αρχιτεκτονικών μοντέλου VGG-19 και ResNet. (Πηγή [129]).....	98
Σχήμα 4.7: Αρχιτεκτονική μοντέλων ResNet50/101/152. (Πηγή [129]) .....	99
Σχήμα 4.8: Αρχιτεκτονική μοντέλου Inception. (Πηγή [130]).....	100
Σχήμα 4.9: Διαδικασία κατά Βάθος Διαχωρίσιμης Συνέλιξης. (Πηγή [132]).....	102
Σχήμα 4.10: Αρχιτεκτονική μοντέλου Xception. (Πηγή [131]) .....	103
Σχήμα 5.1: Συσχέτιση του NMI score με το ποσοστό των παρατηρήσεων που έχουν ομαδοποιηθεί σωστά. (Πηγή [133]) .....	107
Σχήμα 5.2: Διάγραμμα cumulative explained variance για το πρώτο CNN μοντέλο.....	111
Σχήμα 5.3: Αποτελέσματα clustering για το πρώτο CNN μοντέλο.....	112

Σχήμα 5.4: Διάγραμμα cumulative explained variance για το δεύτερο CNN μοντέλο. ....	113
Σχήμα 5.5: Αποτελέσματα clustering για το δεύτερο CNN μοντέλο. ....	114
Σχήμα 5.6: Διάγραμμα cumulative explained variance για το τρίτο CNN μοντέλο.....	115
Σχήμα 5.7: Αποτελέσματα clustering για το τρίτο CNN μοντέλο.....	116
Σχήμα 5.8: Διάγραμμα cumulative explained variance για το VGG16.....	118
Σχήμα 5.9: Αποτελέσματα clustering για το VGG16.....	119
Σχήμα 5.10: Διάγραμμα cumulative explained variance για το VGG19.....	120
Σχήμα 5.11: Αποτελέσματα clustering για το VGG19.....	120
Σχήμα 5.12: Διάγραμμα cumulative explained variance για το ResNet50.....	121
Σχήμα 5.13: Αποτελέσματα clustering για το ResNet50.....	122
Σχήμα 5.14: Διάγραμμα cumulative explained variance για το InceptionV3. ....	123
Σχήμα 5.15: Αποτελέσματα clustering για το InceptionV3. ....	124
Σχήμα 5.16: Διάγραμμα cumulative explained variance για το Xception.....	125
Σχήμα 5.17: Αποτελέσματα clustering για το Xception.....	126
Σχήμα 5.18: Δύο παραλλαγές του χορού «Ο Γιάννης μας αρρώστησε». ....	133
Σχήμα 5.19: Δύο παραλλαγές του χορού «Γκίκνα». ....	134
Σχήμα 5.20: Κινησιογράμματα από τους χορούς Χασάπικος (αριστερά) και Ζωναράδικος (δεξιά).....	135
Σχήμα 5.21: Κινησιογράμματα από τους χορούς «Κουκίτσα» (αριστερά) και «Στης Ματζουράνας» (δεξιά). ....	136
Σχήμα 5.22: Κινησιογράμματα από τους χορούς «Νταουλτζίδικος» (αριστερά) και παραλλαγή του Καρσιλαμά (δεξιά).....	137



## Κατάλογος Πινάκων

---

Πίνακας 5.1: Συγκεντρωτική αξιολόγηση πειραμάτων.....	128
---	-----



# Κεφάλαιο 1

## 1 Εισαγωγή

---

Ο Χορός αποτελεί μια μορφή τέχνης, όπου μέσα από την κίνηση του ανθρώπινου σώματος, άλλοτε με τη συνοδεία μουσικής και άλλοτε μες στη σιωπή, επιτυγχάνεται η έκφραση ιδεών και συναισθημάτων, η επικοινωνία, η μεταβίβαση γνώσεων, ηθών, εθίμων και παραδόσεων, η εκγύμναση και η ψυχαγωγία. Από πολλούς έχει χαρακτηριστεί ως η μητέρα των τεχνών, ανάμεσα σε αυτούς βρίσκεται και ο εθνομουσικολόγος Curt Sachs, ο οποίος υποστηρίζει ότι «η μουσική και η ποίηση υπάρχουν στο χρόνο, η ζωγραφική και η αρχιτεκτονική στο χώρο, όμως ο χορός ζει συγχρόνως στο χρόνο και το χώρο» [1]. Πράγματι, πρόκειται για μια πολυδιάστατη κινητική εμπειρία με αισθητική και εσωτερική αξία, απόλυτα συνυφασμένη με την ανθρώπινη ύπαρξη στην ιστορική, πολιτισμική και κοινωνική της πορεία [2]. Εξάλλου, το ανθρώπινο ον ήδη από τη στιγμή της γέννησής του «χορεύει», καθώς το μοναδικό εργαλείο που απαιτείται για την τέχνη αυτή είναι το ίδιο το ανθρώπινο σώμα, δε χρειάζονται ο λόγος, η γραφή, ούτε άλλα βοηθητικά αντικείμενα για να παραχθεί κίνηση.

Ο χορός επηρεάζεται σημαντικά από τις υπόλοιπες τέχνες, τις καλλιτεχνικές τάσεις, καθώς και τις μεταρρυθμίσεις της κάθε εποχής, αφού ως μορφή τέχνης αφομοιώνει και αναπαράγει τις κοινωνικοπολιτικές και οικονομικές συνθήκες που επικρατούν. Ανά τα χρόνια έχουν αναπτυχθεί πολλά και διαφορετικά είδη χορού, καθένα από τα οποία έχει την δική του τεχνική και ιδιαίτερα χαρακτηριστικά, αντικατοπτρίζοντας την πολιτιστική ταυτότητα των εκάστοτε κοινωνιών. Ανάμεσα σε αυτά, ιδιαίτερος σημαντικό ρόλο παίζει ο Παραδοσιακός Χορός, δεδομένου του ότι αποτελεί κομμάτι πολιτιστικής κληρονομιάς, οπότε η μετάδοσή του στις επόμενες γενιές, που καθιστά δυνατή και τη μελέτη του σαν χορευτικό είδος, είναι ζήτημα υψίστης σημασίας για τη διατήρηση της πολιτισμικής διαφορετικότητας στην εποχή που διανύουμε, αυτή της παγκοσμιοποίησης. Όσον αφορά τον Ελληνικό Παραδοσιακό Χορό, σε συνδυασμό με τη μουσική και το τραγούδι που τον συνοδεύει, αποτελεί αναπόσπαστο κομμάτι του λαϊκού πολιτισμού που εκφράζει την ιδιοσυγκρασία, τα συναισθήματα, τον χαρακτήρα και το πνεύμα του ελληνικού λαού [3], μια ζωντανή κληρονομιά που δεν πρέπει να χαθεί στα βάθη των αιώνων.

Ως εκ τούτου, κρίνεται επιτακτική η ανάγκη διατήρησης αυτού του «άυλου» πολιτισμικού πλούτου, ώστε να διασωθεί από επερχόμενη αλλοίωση ή εξαφάνιση. Και αυτή η διασφάλιση της χορευτικής δημιουργίας μπορεί να γίνει μόνο μέσα από τα γραπτά τεκμήριά της, τη σημειογραφία του χορού. Η σπουδαιότητα των σημειογραφικών συστημάτων για το χορό είναι εμφανής, αν κανείς αναλογιστεί ότι αυτά λειτουργούν κατά αντιστοιχία με τη γραφή για το λόγο, όπως και με τους φθόγγους ή το πεντάγραμμο για τη μουσική [4]. Στην πραγματικότητα πρόκειται για μεθόδους καταγραφής του χορού που χρησιμοποιούν διάφορα σύμβολα, σημεία και σχήματα προκειμένου να αποδώσουν για τον εκάστοτε χορευτικό συνδυασμό πληροφορίες σχετικές με την κατεύθυνση, το επίπεδο, την απόσταση, τη διάρκεια και, ίσως, σπανιότερα και για την ποιότητα της κίνησης.

Το σημειογραφικό σύστημα που μελετάται στην εν λόγω διπλωματική εργασία είναι αυτό που δημιούργησε ο χορογράφος και θεωρητικός του χορού Rudolph von Laban, γνωστό ως Labanotation, το οποίο γνωρίζει παγκόσμια απήχηση και μπορεί να χρησιμοποιηθεί σε όλα τα είδη κίνησης και χορού. Χρησιμοποιεί αφηρημένα σύμβολα που προκύπτουν με βάση τις αρχές του χώρου, της δυναμικής και της ανατομίας του κινούμενου σώματος [4]. Τα σχήματα αυτά οργανώνονται σε έναν τρίγραμμο στύλο, αντίστοιχο του πενταγράμμου της μουσικής, που έχει κάθετο προσανατολισμό και αναπαριστούν τις κινήσεις μεμονωμένων μελών ή και ολόκληρου του σώματος. Η αλληλουχία αυτών των σχημάτων που συνιστούν κάποιο χορευτικό συνδυασμό ορίζεται ως κινησιόγραμμα. Αξίζει να αναφερθεί ότι τόσο η ανάγνωση, όσο και η συγγραφή κινησιογραμμάτων απαιτεί πολύχρονη ενασχόληση με το αντικείμενο αυτό.

Με τη ραγδαία ανάπτυξη της τεχνολογίας, ωστόσο, έχει αλλάξει ο τρόπος με τον οποίο ο χορός αναπαράγεται, μεταδίδεται, καταγράφεται, ακόμη και δημιουργείται. Σε ολόκληρο τον κόσμο υπάρχουν πλέον αρκετές ομάδες ερευνητών που υλοποιούν σημαντικά project, όπως το Wholodance, το WebDANCE και το Terpsichore, σχετικά με την ψηφιοποίηση, την αρχειοθέτηση και τη διάδοση του χορού χρησιμοποιώντας τεχνολογίες αιχμής [5]. Έχουν γίνει, επίσης, προσπάθειες για τη δημιουργία συλλογών σε ψηφιακή μορφή που περιλαμβάνουν βίντεο, εικόνες, κείμενα, κινησιογράμματα, ώστε να είναι ευκολότερη η πρόσβαση και κατ' επέκταση η μελέτη του χορού σε επιστημονική βάση. Τέτοια παραδείγματα αποτελούν ο «Πανδέκτης – Ο κόσμος του ελληνικού χορού» [6] και το «Ερευνητικό πρόγραμμα Θράκη - Μακεδονία» [7], το οποίο χρησιμοποιείται και στα πλαίσια της εργασίας αυτής.

Με τις επιστημονικές αιχμές της Τεχνητής Νοημοσύνης και της Μηχανικής Μάθησης να χαράσσουν τον «τεχνολογικό» πολιτισμό μας, παρατηρούνται ριζικές αλλαγές σε όλους



τους τομείς της ανθρώπινης ζωής και σε όλους τους κλάδους της επιστήμης. Παράλληλα, η δυνατότητα επεξεργασίας του χορού σε ψηφιακή πλέον μορφή δίνει πάτημα για την εφαρμογή αυτών των πρωτότυπων επιτευγμάτων και σε αυτή τη μορφή τέχνης. Μελέτες σχετικά με την ταξινόμηση ειδών χορού [8], [9], την αναγνώριση κινήσεων και χειρονομιών [10], [11], την αυτόματη παραγωγή χορευτικών κινήσεων βάσει του ήχου [12], [13], όλες με χρήση τεχνικών Βαθιάς Μάθησης, έχουν κεντρίσει το ενδιαφέρον της επιστημονικής κοινότητας τις τελευταίες δεκαετίες. Έχει έρθει η στιγμή, λοιπόν, να χρησιμοποιηθούν, ακόμη, για τη διατήρηση του πολιτισμικού πλούτου που κληρονομείται από τον παραδοσιακό χορό, καθώς είναι αδιαμφισβήτητο ότι οι τεχνικές του επιστημονικού αυτού κλάδου μπορούν να δώσουν νέες διαστάσεις στο κομμάτι της σημειογραφίας του χορού, παρά των δυσκολιών του ως αντικείμενο.

## 1.1 Αντικείμενο της Διπλωματικής

Στα πλαίσια αυτής της διπλωματικής εργασίας επιχειρείται η χρήση τεχνικών Βαθιάς Μάθησης και πιο συγκεκριμένα Συνελκτικών Νευρωνικών Δικτύων για το διαχωρισμό σε συστάδες κινησιογραμμάτων που προέρχονται από ελληνικούς παραδοσιακούς χορούς της Θράκης και της Μακεδονίας. Στόχος είναι η δημιουργία ενός συστήματος που θα εκτελεί ομαδοποίηση των καταγεγραμμένων αυτών χορευτικών συνδυασμών, βασιζόμενο στην οπτική ομοιότητά τους, όπως μπορεί να είναι, για παράδειγμα, κάποια μοτίβα που επαναλαμβάνονται, παρόμοιες κινήσεις / στηρίξεις ή ακόμη και ρυθμικές συσχετίσεις.

Είναι γεγονός ότι συγκρίνοντας τα κινησιογράμματα με ανθρώπινο μάτι, μπορεί σχετικά εύκολα κάποιος να παρατηρήσει συγκεκριμένα σύμβολα, τα οποία αναπαριστούν βήματα και συχνά επαναλαμβάνονται πανομοιότυπα ή ελαφρώς τροποποιημένα σε διαφορετικούς χορούς. Στην έρευνά μας, ωστόσο, εκτός των μεμονωμένων συμβόλων, σημασία έχουν και οι ακολουθίες τους, τα μοτίβα, δηλαδή, που σχηματίζονται, οπότε η σύγκριση καθίσταται πιο περίπλοκη για τον εκάστοτε ειδικό. Και φυσικά, όταν το σύνολο δεδομένων περιλαμβάνει μεγάλο αριθμό εικόνων, η αποτελεσματική σύγκρισή τους καταλήγει ανέφικτη. Σε αυτό το σημείο, λοιπόν, οι αλματώδεις εξελίξεις στον τομέα της Μηχανικής Μάθησης, καθώς και της Όρασης υπολογιστών μπορούν να φανούν εξαιρετικά χρήσιμες, προσφέροντας νέες δυνατότητες στην επιστημονική κοινότητα, διαχειριζόμενες το δυσθεώρητο, πολλές φορές, όγκο των πληροφοριών, χωρίς την ανάγκη ανθρώπινης παρέμβασης. Τα Συνελκτικά Νευρωνικά Δίκτυα, ως πρωταθλητές σε θέματα οπτικής ανάγνωσης, ανακαλύπτουν αυτόματα μια ολοένα και πιο σύνθετη ιεραρχία

χαρακτηριστικών, κάτι που τα καθιστά κατάλληλα για την εξαγωγή χρήσιμης πληροφορίας.

Η προτεινόμενη μέθοδος υλοποίησης αφορά στην εφαρμογή τεχνικών μη Επιβλεπόμενης Μάθησης και πιο συγκεκριμένα, μεθόδων Συσταδοποίησης σε ένα σύνολο κινήσιογραμμάτων, τα οποία παρέχονται στο σύστημα που έχει αναπτυχθεί σε μορφή εικόνων. Πριν τη συσταδοποίηση, ωστόσο, ακολουθούνται τα εξής βήματα: αρχικά, γίνεται «χειροκίνητη» επεξεργασία των εικόνων, για τη δημιουργία ενός ομοιογενούς συνόλου δεδομένων, στη συνέχεια, ακολουθεί η εξαγωγή χρήσιμων χαρακτηριστικών τους μέσω οχτώ διαφορετικών Συνελικτικών Νευρωνικών Δικτύων, προ-εκπαιδευμένων και μη, ενώ αμέσως μετά, πραγματοποιείται μείωση της διαστατικότητάς τους, με στόχο την αποτελεσματικότερη και αποδοτικότερη διαχείρισή τους. Τέλος, για την ομαδοποίηση των παρατηρήσεων χρησιμοποιείται διαχωριστικός αλγόριθμος συσταδοποίησης, ο οποίος βασίζεται στα εξαγόμενα χρήσιμα χαρακτηριστικά. Όσον αφορά την αξιολόγηση των αποτελεσμάτων υλοποιείται με μετρικές, οι οποίες συχνά συναντώνται σε τέτοιου είδους προβλήματα και υπολογίζουν την εγκυρότητα της συσταδοποίησης βασιζόμενες σε διάφορα κριτήρια.

Σημαντικό κρίνεται να αναφερθεί πως υπάρχουν λίγες μόνο μελέτες που ασχολούνται με αυτό το ερευνητικό πεδίο, τόσο σε εθνικό, όσο και σε παγκόσμιο επίπεδο λόγω της φύσης του συγκεκριμένου προβλήματος και για αυτό το λόγο έχει μεγάλη σημασία να μελετηθεί η συμπεριφορά και το πώς ανταποκρίνονται αυτές οι τεχνικές σε τέτοια εξειδικευμένα σύνολα δεδομένων που χρήζουν όχι μόνο ερευνητικής, αλλά και ιστορικής, ανθρωπολογικής και πολιτισμικής αξίας.

## 1.2 Κίνητρο της Διπλωματικής

Το κίνητρο για την επιλογή του συγκεκριμένου θέματος και τη συγγραφή της εν λόγω εργασίας πηγάζει από τη βαθιά μου αγάπη για την τέχνη του χορού και από το ενδιαφέρον μου να βρω τρόπους να συνδυάσω δύο εντελώς διαφορετικές επιστήμες που έχουν επηρεάσει σημαντικά τη ζωή μου. Όλα ξεκίνησαν σε ένα μάθημα χορού και, μάλιστα, παραδοσιακού, όπου η καθηγήτρια αναφέρθηκε στην ύπαρξη ενός συστήματος καταγραφής της κίνησης που περιλαμβάνει σύμβολα και αποτελεί ένα είδος κωδικοποιημένης «γλώσσας» για το χορό. Παρόλα αυτά, μας τόνισε πως είναι ένα αρκετά δύσκολο αντικείμενο και χρειάζεται χρόνο και ειδική εκπαίδευση, ώστε να μπορέσει κάποιος να διαβάσει και αντιστοίχως να γράψει τον «χορό». Τότε, σκέφτηκα ότι, ίσως θα

ήταν ενδιαφέρον και θα διευκόλυνε σημαντικά την κοινότητα των επιστημόνων, αλλά και των θεωρητικών του χορού, εάν αυτό το σύστημα γραφής μπορούσε να αποκωδικοποιηθεί με τη βοήθεια της τεχνητής νοημοσύνης και αυτό αποτέλεσε το έναυσμα για περαιτέρω έρευνα από πλευράς μου. Έτσι, μετά από μελέτη κατέληξα ότι αποτελεί μια καλή ευκαιρία να ασχοληθώ με ένα θέμα που βασίζεται στη Μηχανική Μάθηση, ένα πεδίο που μου κίνησε σε μεγάλο βαθμό το ενδιαφέρον κατά τη διάρκεια των σπουδών μου, αλλά συγχρόνως έχει και καλλιτεχνικές και πολιτισμικές προεκτάσεις, πράγμα το οποίο μου έδωσε επιπλέον κίνητρο να αφοσιωθώ στην εργασία αυτή.

### 1.3 Παρεμφερείς Μελέτες

Ο χορός, δυστυχώς, αποτελεί μία από τις λιγότερο «ακαδημαϊκά μελετημένες» τέχνες, ίσως λόγω του παραστατικού χαρακτήρα του και της δυσκολίας του ως προς την έρευνα σε θεωρητικό επίπεδο. Αναφορικά με το κομμάτι της σημειογραφίας της κίνησης, αποτελεί, επίσης, ένα πεδίο που αν και τόσο χρήσιμο, δεν έχει λάβει την αναγνώριση που του αρμόζει από την επιστημονική κοινότητα. Παρόλα αυτά, προσφάτως έχουν γίνει σημαντικές προσπάθειες από αρκετούς ερευνητές, οι οποίοι χρησιμοποιούν πρωτοπόρα τεχνολογικά μέσα, ώστε να συμπεριλάβουν την τέχνη του χορού, να την αναλύσουν, να την καταγράψουν, να την ψηφιοποιήσουν και εν τέλει να την «αναβαθμίσουν». Πιο συγκεκριμένα, το σύστημα καταγραφής Labanotation συναντά τα τελευταία χρόνια την Τεχνολογία και τα αποτελέσματα είναι ελπιδοφόρα.

Αρχικά, σημαντική είναι η συνεισφορά των ερευνητών που δημιούργησαν λογισμικά για την ψηφιακή καταγραφή των κινησιογραμμάτων Laban. Ανάμεσα σε αυτά το διασημότερο είναι το Laban Writer [14], το οποίο, μάλιστα, προωθεί και το Διεθνές Συμβούλιο Κινησιογραφίας Laban (International Council of Kinetography Laban -ICKL). Παρόμοια λογισμικά που έχουν αναπτυχθεί για τη διευκόλυνση της καταγραφής του χορού σε ηλεκτρονικό υπολογιστή είναι το Labanatory [15], το Calaban [16], καθώς και το LED&LINTEL [17]. Φυσικά, είναι πολύ πιο βολική και αποδοτικότερη χρονικά η ψηφιακή καταγραφή των dance scores, συγκριτικά με την καταγραφή τους στο χαρτί, ενώ, επιπλέον, είναι και πιο ομοιόμορφη, δεδομένου του ότι ο τρίγραμμος στύλος, τα σύμβολα, τα σχήματα, οι κατευθύνσεις έχουν συγκεκριμένη μορφή και δε σχεδιάζονται από τον κάθε επαγγελματία με διαφορετικό τρόπο, κάτι που δυσχεραίνει ακόμη περισσότερο τη μελέτη και τη σύγκρισή τους.

Επίσης, έχουν αναπτυχθεί λογισμικά που επιτρέπουν τη μετατροπή των κινησιογραμμάτων σε τρισδιάστατη αναπαράσταση (3D human animation), διευκολύνοντας σημαντικά την «ανάγνωσή» τους, μειώνοντας το χρόνο που απαιτείται για την αποκωδικοποίησή τους και συμβάλλοντας ενδεχομένως σε εκπαιδευτικούς σκοπούς, εφόσον κάτι τέτοιο θα μπορούσε να χρησιμοποιηθεί ακόμη και για εκμάθηση χορογραφιών που είναι καταγεγραμμένες σε αυτή τη μορφή. Τέτοια ερευνητικά έργα αποτελούν ο Laban Editor [18], ο Laban Dancer [19] και το Life Forms [20].

Άλλοι ερευνητές έχουν στρέψει το ενδιαφέρον τους στην ανάπτυξη μιας «χορευτικής» γλώσσας που είναι κατανοητή τόσο από τον άνθρωπο, όσο και από τις μηχανές. Κίνητρό τους υπήρξε το γεγονός ότι τα προαναφερθέντα λογισμικά, πέραν της ψηφιακής σχεδίασης και 3D αναπαράστασης των κινησιογραμμάτων, δε δίνουν τη δυνατότητα για ερμηνεία και ανάλυση των κινήσεων, για εύρεση συγκεκριμένων κινητικών μοτίβων, ούτε για κάποιου είδους γραπτή αναπαράσταση των επιμέρους συμβόλων που χρησιμοποιούνται. Έτσι, δημιουργήθηκαν η LabanXML [21], η MovementXML [22], καθώς και άλλες πρωτότυπες αναπαραστάσεις, όπως με τη βοήθεια γλωσσών οντολογιών [23], με σκοπό να συμπεριλάβουν τη σημασιολογία του Labanotation.

Τέλος, ένα πολύ σημαντικό βήμα έχει γίνει στα πλαίσια της αυτόματης παραγωγής κινησιογραμμάτων. Ορισμένες τεχνικές εστιάζουν στην χωρική ανάλυση των κινήσεων [24]–[26] για την αυτόματη εξαγωγή σημειογραφίας. Βασίζονται σε ορισμένους κανόνες, ώστε μετά την κατάτμηση της κίνησης (motion segmentation) να γίνει η κατάλληλη αντιστοίχιση των βασικών κινήσεων που προκύπτουν σε σύμβολα Laban. Οι Hachimura και Nakamura [27] ήταν οι πρώτοι που πρότειναν την παραγωγή Labanotation, μέσω δεδομένων που προέκυψαν από συστήματα σύλληψης της κίνησης (motion captured data). Αργότερα, δημιουργήθηκε το GenLaban [28], το οποίο όμως δε λάμβανε υπόψη του το ρυθμικό περιεχόμενο των χορογραφιών όταν δημιουργούσε κινησιογράμματα. Τα συστήματα αυτά που βασίζονταν σε κανόνες (rule-based methods), δεν κατάφεραν να καταγράψουν απολύτως σωστά την ανθρώπινη κίνηση, η οποία είναι τόσο περίπλοκη και απρόβλεπτη που δεν μπορεί να περιοριστεί σε κανόνες. Πιο επιτυχημένες προσπάθειες έγιναν προσφάτως από τους Zhou et al. [29] και Li et al. [30]. Οι τελευταίοι πρότειναν μια μέθοδο μοντελοποίησης της ανθρώπινης κίνησης βασισμένη σε μαρκοβιανές αλυσίδες (Hidden Markov chains). Ακόμη, άλλοι ερευνητές αξιοποιούν τα πλεονεκτήματα της μηχανικής μάθησης και των Νευρωνικών Δικτύων [31], [32], αποφέροντας βελτιωμένα αποτελέσματα. Φυσικά, αυτές αποτελούν μερικές μόνο από τις σημαντικότερες μελέτες σχετικά με την αυτόματη παραγωγή κινησιογραμμάτων. Το πεδίο αυτό μόλις που έχει κάνει την εμφάνισή του και έχει πολλά ακόμη να προσφέρει.

## 1.4 Οργάνωση της Διπλωματικής

Η παρούσα εργασία χωρίζεται σε τρία κύρια μέρη, το Θεωρητικό, το Πρακτικό και τον Επίλογο. Καθένα από αυτά περιλαμβάνει επιμέρους κεφάλαια, τα οποία συνολικά προκύπτουν έξι. Το συγκεκριμένο κεφάλαιο αποτελεί μια εισαγωγική παρουσίαση της εργασίας, προοίμιο για τα όσα έχουν γραφεί στο υπόλοιπο του εγγράφου.

Το Θεωρητικό μέρος περιλαμβάνει το δεύτερο και τρίτο κεφάλαιο. Στο Κεφάλαιο 2, περιγράφεται η έννοια του χορού, καθώς και της σημειογραφίας του, δίνοντας έμφαση στο σημειογραφικό σύστημα Labanotation, ενώ επίσης γίνεται σύνδεσή του με την τεχνολογία και το πώς τα δύο αυτά πεδία αλληλεπιδρούν. Στο Κεφάλαιο 3, δίδεται το θεωρητικό υπόβαθρο που αφορά τη Μηχανική Μάθηση, εστιάζοντας στη μέθοδο της Συσταδοποίησης, καθώς και στα Συνελκτικά Νευρωνικά Δίκτυα που χρησιμοποιήθηκαν στη συνέχεια.

Το Πρακτικό μέρος περιλαμβάνει το τέταρτο και πέμπτο κεφάλαιο, όπου περιγράφεται η πρακτική προσέγγιση του θέματος της εργασίας. Πιο συγκεκριμένα, στο Κεφάλαιο 4 παρουσιάζεται το σύνολο δεδομένων, τα εργαλεία, καθώς και οι αρχιτεκτονικές των δικτύων που χρησιμοποιήθηκαν για την εκπόνηση της πειραματικής διαδικασίας. Το Κεφάλαιο 5 είναι αφιερωμένο στην παρουσίαση των αποτελεσμάτων που προέκυψαν, χρησιμοποιώντας διαφορετικές μετρικές αξιολόγησης και ακολουθεί σχολιασμός τους.

Ο Επίλογος περιλαμβάνει το έκτο και τελευταίο κεφάλαιο της εργασίας. Σε αυτό γίνεται η σύνοψη του εγγράφου, παρουσιάζονται τα συμπεράσματα, ενώ επίσης προτείνονται πιθανές μελλοντικές επεκτάσεις επί του θέματος.



# Μέρος Ι

Θεωρητικό Μέρος

---





## Κεφάλαιο 2

### 2 Η τέχνη του Χορού

---

Χορός είναι μια μορφή τέχνης, η τέχνη της κίνησης του σώματος [33]. Είναι το αποτέλεσμα μιας μοναδικής εσωτερικής κίνησης, μιας παρόρμησης που, με τη συνεργασία του συναισθήματος και της σκέψης, οδηγεί σε μια βαθιά σημασία. Τότε είναι που η κίνηση αποκτά νόημα, μεταδίδοντας μια αλήθεια, μια προσωπική ανακάλυψη [34]. Πράγματι, κάθε μικρή ή μεγάλη κίνηση εκφράζει κάτι, μια σκέψη, ένα συναίσθημα, μια ιδέα.

Έχουν δοθεί πολλοί και ενδιαφέροντες ορισμοί ανά τους αιώνες σχετικά με την τέχνη αυτή και θα ήθελα να αναφερθώ σε δύο εξ αυτών που παρουσιάζουν μεγάλο ενδιαφέρον. Σύμφωνα με τον Πλάτωνα, ο χορός είναι «η ενστικτώδης επιθυμία να ερμηνευθούν οι λόγοι, δια των κινήσεων, ολοκλήρου του σώματος» και εκφράζει την πεποίθηση ότι ο «α-χόρευτος» άνθρωπος είναι επί της ουσίας ακαλλιέργητος. Επίσης, σημαντικό κρίνω να αναφερθεί ο ορισμός που έδωσε η Martha Graham, μια από τις σημαντικότερες εκπροσώπους του μοντέρνου χορού, χορεύτρια και χορογράφος, που έπαιξε καθοριστικό ρόλο στην πορεία και την εξέλιξή του. Για εκείνη, «ο χορός είναι η κρυμμένη γλώσσα της ψυχής».

Αυτό που προσπάθησαν να εκφράσουν πολλοί, μεγάλοι καλλιτέχνες είναι ότι ο χορός αποτελεί τρόπο έκφρασης, ένα μέσο επικοινωνίας χορευτή – θεατή, με τον πρώτο να χρησιμοποιεί το χρόνο, το χώρο, τις διαφορετικές ποιότητες (απαλή, έντονη, ελεγχόμενη, απότομη, οδηγημένη κίνηση κλπ.) ή ακόμη και τις ποικίλες θέσεις που μπορεί να πάρει το ανθρώπινο σώμα, ώστε να μεταδώσει το μήνυμα που επιθυμεί.

#### 2.1 Ιστορική αναδρομή

Ο χορός έχει τις ρίζες του στις πρώτες ήδη κοινωνίες των ανθρώπων. Προτού ανακαλυφθεί ο προφορικός λόγος ή η γραφή, ο άνθρωπος χρησιμοποίησε την κίνηση για να εκφράσει τον υλικό, ψυχικό και πνευματικό του κόσμο. Έτσι, ο χορός χρησιμοποιήθηκε σαν μέσο επικοινωνίας και διαπαιδαγώγησης, συστατικό στοιχείο των πρωτόγονων κοινωνιών. Μέχρι και την Κλασική Εποχή, θεωρείται μια σεβαστή μορφή τέχνης,

διαδραματίζοντας σημαντικό ρόλο στα πλαίσια της κοινωνίας. Ωστόσο, κατά την εποχή του Μεσαίωνα με την εγκαθίδρυση του Χριστιανισμού και την Εκκλησία να διαχωρίζει την ψυχή από το σώμα και να καταδικάζει απερίφραστα κάθε πρακτική που χρησιμοποιεί το σώμα ως μέσο ευχαρίστησης, όπως είναι ο χορός, η τέχνη αυτή βιώνει την παρακμή.

Κατά τον 17<sup>ο</sup> και 18<sup>ο</sup> αιώνα, εμφανίζεται το μπαλέτο, το οποίο γνωρίζει μεγάλη άνθιση, ως ψυχαγωγικό μέσο της Αυλής. Σταδιακά, γίνεται προσιτό σε ένα ευρύτερο κοινό, συστηματικοποιείται μέσω της τεχνικής, ενώ επίσης γίνεται αντικείμενο θεωρητικής μελέτης. Ο 19<sup>ος</sup> αιώνας, ωστόσο, είναι αυτός που χαρακτηρίζεται ως χρυσός αιώνας του μπαλέτου, κατά τον οποίο φτάνει στο απόγειό του με τις μοναδικές χορογραφίες του Marius Petipa. Την εποχή εκείνη μιλάμε πια για κλασικό μπαλέτο, το οποίο έχει συνδεθεί με την τεχνική αρτιότητα, την αρμονία και τη συμμετρία.

Στα χρόνια που έπονται κάνει την εμφάνισή του ο μοντέρνος χορός σαν αντίδραση στον ακαδημαϊσμό και την κλασική φόρμα του μπαλέτου. Πρόκειται για την περίοδο του μεταμοντέρνου χορού που αλλάζει δυναμικά το περιεχόμενο και τη μορφή της κίνησης, ακολουθώντας τις επιταγές του τότε κοινωνικοπολιτικού πλαισίου και υπηρετώντας σημαντικά κινήματα της εποχής, όπως αυτά που αφορούσαν την ισότητα και την ειρήνη [35].

Αργότερα, κατά τη δεκαετία του 1970 – 1980, η ανάγκη για περαιτέρω εξερεύνηση της κίνησης του ανθρώπινου σώματος, οδήγησαν στην ανάδυση της τεχνικής του σύγχρονου χορού. Έχει επιρροές από όλες τις προηγούμενες χορευτικές τεχνικές, αλλά και από το θέατρο, την ακροβασία και το τσίρκο, από διάφορα αθλήματα και πολεμικές τέχνες.



Σχήμα 2.1: Στιγμιότυπο από την παράσταση *Grand Finale* - Hofesh Shechter Company (2018). (Πηγή

[36])

Το σύστημα χορού που δημιουργείται δεν είναι πλέον αυστηρά οριοθετημένο και τεχνικό, αλλά αποτελεί ανοιχτό πεδίο έρευνας της ανθρώπινης κίνησης που συνεχώς εξελίσσεται [35]. Πρόκειται για μια τεχνική που δίνει χώρο στην προσωπική κινητική έκφραση και σέβεται την αρχή της εξατομίκευσης. Στη σύγχρονη, λοιπόν, εποχή, ο χορός ως παραστατική τέχνη, εκφράζει σκέψεις, συναισθήματα και ιδέες που ο/η χορογράφος και ο/η χορευτής προσπαθούν να μεταδώσουν στο κοινό τους, αλλά ταυτοχρόνως αποτελεί για τον καθένα, επαγγελματία ή ερασιτέχνη, ευκαιρία για απελευθέρωση, διοχέτευση της ενέργειας και έκφραση μέσω της προσωπικής κίνησης.

Συμπερασματικά, ο χορός επηρεάζεται σημαντικά από τις υπόλοιπες τέχνες, τις καλλιτεχνικές τάσεις, καθώς και τις κοινωνικοπολιτικές μεταρρυθμίσεις της κάθε εποχής. Ανά τα χρόνια, έχουν δημιουργηθεί πολλά και διαφορετικά είδη χορού, ορισμένα εκ των οποίων έχουν διατηρήσει την ισχύ τους κι άλλα, τα οποία σταδιακά εκλείπουν. Ανάμεσα σε αυτά, ιδιαίτερος σημαντικό ρόλο παίζει ο Παραδοσιακός Χορός, καθώς αντανakλά τα διαφορετικά ήθη και έθιμα, θρησκείες και ιδιαιτερότητες της κάθε κοινωνίας και με αυτόν τον τρόπο, αποτελεί μέσο μεταβίβασης της πολιτιστικής κληρονομιάς από γενιά σε γενιά.

## 2.2 Ελληνικός Παραδοσιακός Χορός

Αποτελεί ένα είδος χορού που αναπτύχθηκε στον ελλαδικό χώρο και εκφράζει την ιδιοσυγκρασία, τα συναισθήματα, τον χαρακτήρα, το πνεύμα του ελληνικού λαού [3] και είναι άμεσα συνυφασμένος με τον πολιτισμό και την ιστορία του ελληνισμού. Είναι αποτέλεσμα της πολυμορφίας των λειτουργιών της ζωής του Έλληνα, ποικίλει από τόπο σε τόπο και αποτελεί δημιούργημα λαού με μακρόχρονη πολιτισμική παράδοση [37], μια κληρονομιά που μεταφέρεται από γενιά σε γενιά. Έτσι, η μορφή του έχει να κάνει με την ιστορία της εκάστοτε περιοχής, την παράδοση, τις κλιματολογικές συνθήκες, τη μουσική, ακόμη και την ενδυμασία.

Η διαμόρφωση των χορών αυτών αντικατοπτρίζει το συμμετοχικό, κοινωνικό και ομαδικό τους χαρακτήρα, δημιουργώντας σχέσεις και αλληλεπιδράσεις μεταξύ των ατόμων. Βασικό χαρακτηριστικό τους είναι το «ανώνυμο και συλλογικό» [38]. Ειδικότερα κατά την περίοδο της Τουρκοκρατίας, όπου ο χορός αποτελούσε ένα ισχυρό όπλο ενάντια στην αλλοτρίωση και την απώλεια της εθνικής συνείδησης, παρατηρείται άνθιση του λαϊκού χορού. Νέα είδη χορών γεννιούνται, τα οποία συχνά περιγράφουν ηρωικά κατορθώματα και σημαντικές στιγμές της Ελλάδας. Μετά την απελευθέρωση και τη συγκρότηση του πρώτου ελληνικού κράτους, υπήρξε στροφή του ενδιαφέροντος των

λογοτεχνών προς την εθνική μας παράδοση και κατά συνέπεια και τους παραδοσιακούς χορούς. Ωστόσο, η αθηναϊκή κοινωνία στρέφεται σταδιακά σε ευρωπαϊκά πρότυπα, ενώ αρνητική επίδραση στη διάδοσή τους έχει και ο Β' Παγκόσμιος Πόλεμος, η βιομηχανική επανάσταση, όπως και η αστικοποίηση που ερημώνουν την ύπαιθρο με τον αγροτικό πληθυσμό, ο οποίος αποτελεί την καρδιά του λαϊκού πολιτισμού, να μαραζώνει [39]. Παρόλα αυτά, ο ελληνικός παραδοσιακός χορός εξακολουθεί να αποτελεί αναπόσπαστο κομμάτι του λαϊκού πολιτισμού που συνδέεται στενά με τη μουσική και το τραγούδι. Θεωρείται σημαντικό στοιχείο πολιτισμικής συνέχειας και παράδοσης του ελληνικού λαού γι' αυτό και γίνονται προσπάθειες για τη διδασκαλία του σε όλες τις βαθμίδες της εκπαίδευσης.

Οι παραδοσιακοί χοροί παρουσιάζουν μεγάλη ποικιλία στον ελλαδικό χώρο. Γι' αυτό το λόγο συχνά ταξινομούνται σε διάφορες κατηγορίες ανάλογα με το ρυθμό (δίσημος, τρίσημος, τετράσημος κ.ο.κ.), το θέμα (πολεμικοί, θρησκευτικοί χοροί κλπ.), το σχήμα (κυκλικοί, αντικριστοί) ή και τον τόπο, όπου χορεύονται. Πράγματι, κάθε πολιτισμική κοινότητα της Ελλάδας έχει τους δικούς της τοπικούς χορούς, οι οποίοι διαφέρουν από πόλη σε πόλη, ακόμη και από χωριό σε χωριό. Με τον όρο πολιτισμική κοινότητα εννοούμε ομάδες ανθρώπων που έχουν κοινή καταγωγή και συνδέονται με δεσμούς κοινής πολιτισμικής δράσης, όπως συναντώνται στη Θράκη, τη Μακεδονία, την Ήπειρο, τη Θεσσαλία, τη Ρούμελη, το Μοριά, την Κρήτη, τα νησιά του Ιονίου και του Αιγαίου, καθώς και στην Κύπρο. Όσον αφορά την ονομασία των χορών αυτών, συνήθως προέρχεται από



*Σχήμα 2.2: Συγκαθιστός ή Μαντηλάτος – Παραδοσιακός Χορός Θράκης. (Πηγή [40])*

τον τρόπο με τον οποίο χορεύεται ο καθένας ή από τη γεωγραφική περιοχή στην οποία έχει πρωτοχορευτεί. Στην παρούσα διπλωματική εργασία μελετώνται παραδοσιακοί χοροί της Θράκης και της Μακεδονίας, ανάμεσα στους οποίους ξεχωρίζουν ο Ζωναράδικος, ο Χασάπικος, η Μπαϊντούσκα, ο Συγκαθιστός και ο Καρσιλαμάς.

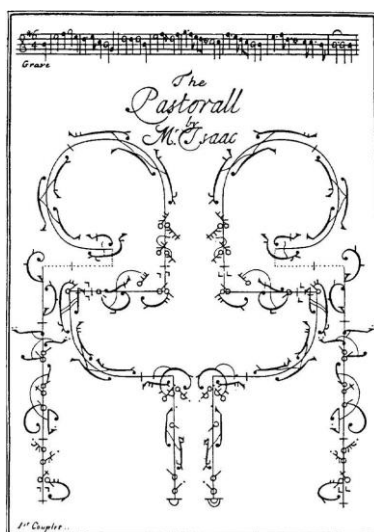
Συνολικά, ο ελληνικός παραδοσιακός χορός, ως μέσο έκφρασης και ως μορφή τέχνης, ως αντικείμενο με γνωστικό και κιναισθητικό περιεχόμενο, αποτελεί βασικό δομικό συστατικό στοιχείο της νεότερης πολιτιστικής κληρονομιάς [41]–[43] και, ειδικότερα, της αποκαλούμενης άυλης πολιτιστικής κληρονομιάς (Intangible Cultural Heritage - ICH), η οποία τελεί υπό την προστασία της UNESCO, διαφυλάσσοντας τη συνεχή δημιουργικότητα της ανθρώπινης σκέψης [2].

## 2.3 Σημειογραφία του Χορού

Η λέξη «σημειογραφία», ως ακαδημαϊκός όρος της επιστήμης του χορού, είναι ένα είδος παραστατικής γραφής, η οποία χρησιμοποιώντας σημεία και σύμβολα, περιγράφει στο χαρτί το αποτύπωμα της κίνησης που επιτελεί το ανθρώπινο σώμα [44]. Πρόκειται για ένα θεμελιώδες εργαλείο, με πολιτισμικά και εκπαιδευτικά οφέλη, απαραίτητο για την έρευνα και τη μελέτη του χορού, καθώς δίνει τη δυνατότητα για διατήρηση και ίσως αναβίωση χορογραφιών [4], που δε θα μπορούσαν να έχουν διασωθεί με άλλο τρόπο. Πράγματι, η καταγραφή της χορευτικής κίνησης διασφαλίζει τη διάδοση της χορευτικής γνώσης, κάτι το οποίο δεν είναι πάντοτε δυνατό μόνο μέσω του προφορικού λόγου ή ακόμη και μέσω ενός βίντεο. Επιπλέον, επιτρέπει την ανάλυση και την ταξινόμηση των χορών, δίνοντας πρόσβαση σε διαφορετικούς χορευτικούς πολιτισμούς, την ομαδοποίηση και τη σύγκρισή τους, τη συστηματοποίησή τους, θεμελιώνοντας την ίδια την επιστήμη του χορού. Επιπροσθέτως, κάτι εξαιρετικά ενδιαφέρον είναι το γεγονός ότι κάνει τον χορό προσβάσιμο σε ένα πολύ πιο ευρύ κοινό, ακόμη και σε ανθρώπους με οπτικά ή ακουστικά προβλήματα, παρέχοντας πολύτιμο υλικό προς διερεύνηση. Με λίγα λόγια, αποτελεί ένα πολύ σημαντικό ερευνητικό πεδίο και δεδομένου του ότι η ιστορία του ανθρώπου ξεκινά με τα πρώτα γραπτά τεκμήρια [4], η καταγραφή του χορού αποτελεί ένα πέρασμα από «το μύθο στην ιστορία», όπως υποστηρίζει και η Doris Humphrey [45].

Απόπειρες για καταγραφή της κίνησης και του χορού υπήρξαν από τους αρχαίους χρόνους, όμως οι πρώτες ουσιαστικές προσπάθειες εντοπίζονται κατά τον 15<sup>ο</sup> αιώνα. Έκτοτε, έχουν αναπτυχθεί περισσότερα από 85 σημειογραφικά συστήματα μόνο στο δυτικό κόσμο. Ανάμεσα σε αυτά υπάρχουν συστήματα, τα οποία χαρακτηρίζονται από

γράμματα, λέξεις ή συντμήσεις λέξεων, ενώ άλλα περιλαμβάνουν οπτικά γραφήματα, όπως σύμβολα για το χώρο ή αφηρημένα σχήματα που ενίοτε αναπτύσσονται παράλληλα με τη μουσική παρτιτούρα [44].



Σχήμα 2.3: Παράδειγμα καταγραφής χορού με το σύστημα του Feuillet. (Πηγή [45])

[45] Το σύστημα σημειογραφίας της χορευτικής κίνησης που επικράτησε, ωστόσο, είναι αυτό που δημιούργησε ο Rudolph von Laban, Ούγγρος χορογράφος, παιδαγωγός και θεωρητικός της κίνησης και του ρυθμού. Το ονόμασε *Labanotation* και αποτελεί μια εξειδικευμένη γνώση, μια κωδικοποιημένη γλώσσα οικουμενικής εμβέλειας που περιλαμβάνει μεγάλη ποικιλία σημείων και συμβόλων, από την οποία κάθε φορά επιλέγονται εκείνα που εξυπηρετούν το εκάστοτε είδος χορού. Στηρίζεται στην ανατομία του ανθρώπινου σώματος, είναι ευέλικτο και προσαρμοστικό, συνοπτικό και συγχρόνως δυναμικό, βρίσκοντας πολλαπλές εφαρμογές και επίπεδα ανάλυσης [44]. Πρόκειται για ένα σύστημα καταγραφής που σφράγισε τη μελέτη του χορού σε επιστημονική βάση.

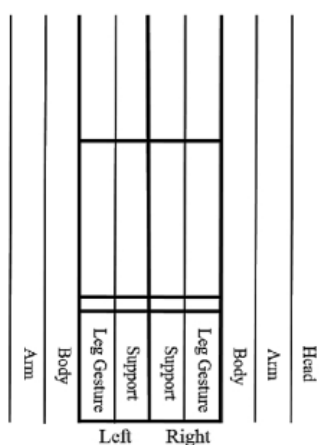
Στην Ελλάδα, η χρήση σημειογραφικών συστημάτων είναι σχετικά πρόσφατη. Πιο συγκεκριμένα, η ιστορία της συστηματικής έρευνας του παραδοσιακού χορού χρονολογείται στις αρχές του 20<sup>ου</sup> αιώνα, μέσω κυρίως λεκτικών περιγραφών. Από τη δεκαετία του 1960 – 1990, η καταγραφή των χορών με τη βοήθεια κινηματογραφικού φιλμ αποτελούσε το κύριο μέσο αποτύπωσης της κίνησης, ενώ παράλληλα με αυτό ερευνητές στα πλαίσια των διδακτορικών τους διατριβών ή ακόμη σε άρθρα και δισκογραφικές παραγωγές παρουσιάζουν γραφήματα και παρτιτούρες χορού με δύο κυρίαρχα συστήματα, του Laban και του Conté [44].

### 2.3.1 Labanotation

Η κινησιογραφία του *Laban* αποτελεί ένα σύστημα καταγραφής του χορού που καθόρισε την επιστημονική εξέλιξη της τέχνης αυτής. Σε αυτή γίνεται χρήση διαφόρων συμβόλων και σχημάτων, έτσι ώστε να αποτυπωθούν οι 4 διαστάσεις του χορού, όπου 4η θεωρείται ο χρόνος, στις 2 διαστάσεις του χαρτιού. Πρόκειται για έναν είδος κοινής γλώσσας, το οποίο, όμως, θεωρείται συχνά κάπως «ιδιάζον» και απαιτείται εξοικείωση, τόσο για να μπορέσει κάποιος να αναγνωρίσει τα σύμβολα και κατ' επέκταση την ακολουθία των χορευτικών κινήσεων, όσο και για να μπορέσει να καταγράψει ο ίδιος το χορευτικό συνδυασμό που επιθυμεί. Αναλύει και αποτυπώνει στο χαρτί διαφορετικής ποιότητας κινήσεις, εφόσον το ενδιαφέρον της καταγραφής της κίνησης εστιάζει στην ίδια την κίνηση και στο κινούμενο σώμα, ενώ δεν ταυτίζεται με συγκεκριμένο είδος χορού [44].

Να αναφερθεί, αρχικά, ότι η καταγραφή μιας οποιασδήποτε αλληλουχίας κινήσεων καλείται κινησιόγραμμα ή αλλιώς κινητόγραμμα (kinetogram), ενώ η καταγραφή μιας χορογραφίας, δηλαδή μιας πλήρους χορογραφικής σύνθεσης, ονομάζεται παρτιτούρα χορού (dance score) [4].

Η σημειογραφία της χορευτικής κίνησης του Laban αποτυπώνεται σε ένα στύλο τριών κάθετων γραμμών, γνωστό ως τρίγραμμο στύλο. Σαν βάση για το κινησιόγραμμα χρησιμοποιήθηκε το πεντάγραμμο της μουσικής, περιστρεμμένο κατά 90° και από το οποίο διατηρήθηκαν μόνο η πρώτη, η τρίτη και η πέμπτη γραμμή. Αυτές οι τρεις γραμμές αναπαριστούν το ανθρώπινο σώμα, με την κεντρική να αντιπροσωπεύει τον κεντρικό άξονά του, διαχωρίζοντάς το στα δύο ημιμόρια, το δεξί και το αριστερό. Έτσι, ό,τι σημειώνεται στα δεξιά αφορά τις κινήσεις του δεξιού ημιμορίου και αντίστοιχα για την αριστερή πλευρά. Πέρα από αυτές, ωστόσο, υπάρχουν και νοητές γραμμές. Τα διαστήματα

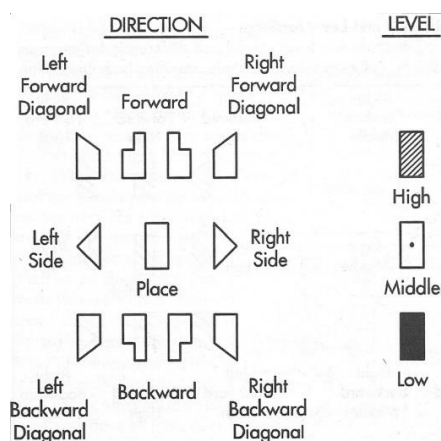


Σχήμα 2.4: Ονοματοθεσία στηλών – με έντονο χρώμα απεικονίζονται οι τρεις βασικές γραμμές του τρίγραμμου στύλου, ενώ οι υπόλοιπες είναι νοητές. (Πηγή [46])

[46]μεταξύ υπαρκτών και νοητών γραμμών ονομάζονται στήλες και αντιστοιχούν η καθεμία σε συγκεκριμένο μέρος του σώματος, δηλαδή στα άνω και κάτω άκρα, τον κορμό και το κεφάλι [4].

Ανάμεσα στις στήλες τοποθετούνται ορισμένα σύμβολα, τα οποία περιλαμβάνουν διαφορετικά σχήματα, σε διάφορα μεγέθη και σκιαγραφήσεις, καθορίζοντας την κατεύθυνση, το επίπεδο και τη διάρκεια των εκάστοτε στηρίξεων / κινήσεων. Τόσο για τη στήριξη, όσο και για την κίνηση του σώματος χρησιμοποιείται ως βασικό σχήμα το ορθογώνιο παραλληλόγραμμο [4]. (Ο Laban θεωρεί ως στήριξη εκείνη που φέρει το βάρος του σώματος, ενώ η κίνηση όχι.) Πιο συγκεκριμένα, ανάλογα με:

- Το σχήμα αποτυπώνεται η κατεύθυνση.
- Την σκιαγράφηση αποτυπώνεται το επίπεδο.
- Τη θέση τους μέσα στο κινήσιόγραμμα αποτυπώνεται ο χρόνος που γίνεται η κίνηση, όσο πιο χαμηλά βρίσκονται τόσο νωρίτερα εκτελούνται οι εν λόγω κινήσεις, εφόσον διαβάζεται πάντοτε από κάτω προς τα πάνω.
- Το μέγεθος αποτυπώνεται η διάρκεια της κίνησης, όσο μεγαλύτερα, τόσο μεγαλύτερη και η διάρκεια.
- Την τοποθέτηση σε μία από τις στήλες αποτυπώνεται ποιο μέρος του σώματος πραγματοποιεί την κίνηση [44].



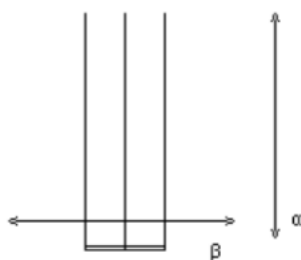
Σχήμα 2.5: Σύμβολα που καθορίζουν την κατεύθυνση και το επίπεδο της κίνησης. (Πηγή [47])

[47]Φυσικά, υπάρχουν και άλλα σύμβολα για την αναπαράσταση συγκεκριμένων σημείων του σώματος, τη μέτρηση του χώρου (απόστασης), τις στροφές και τη φορά τους, τη μετακίνηση στο χώρο, αλλά και ορισμένα βασικά και συχνά χρησιμοποιούμενα «πακέτα» συμβόλων για την απεικόνιση βασικών λαβών, ιδίως στον ελληνικό παραδοσιακό χορό, διαφόρων ειδών αλμάτων, καθώς και σύμβολα που καθορίζουν την



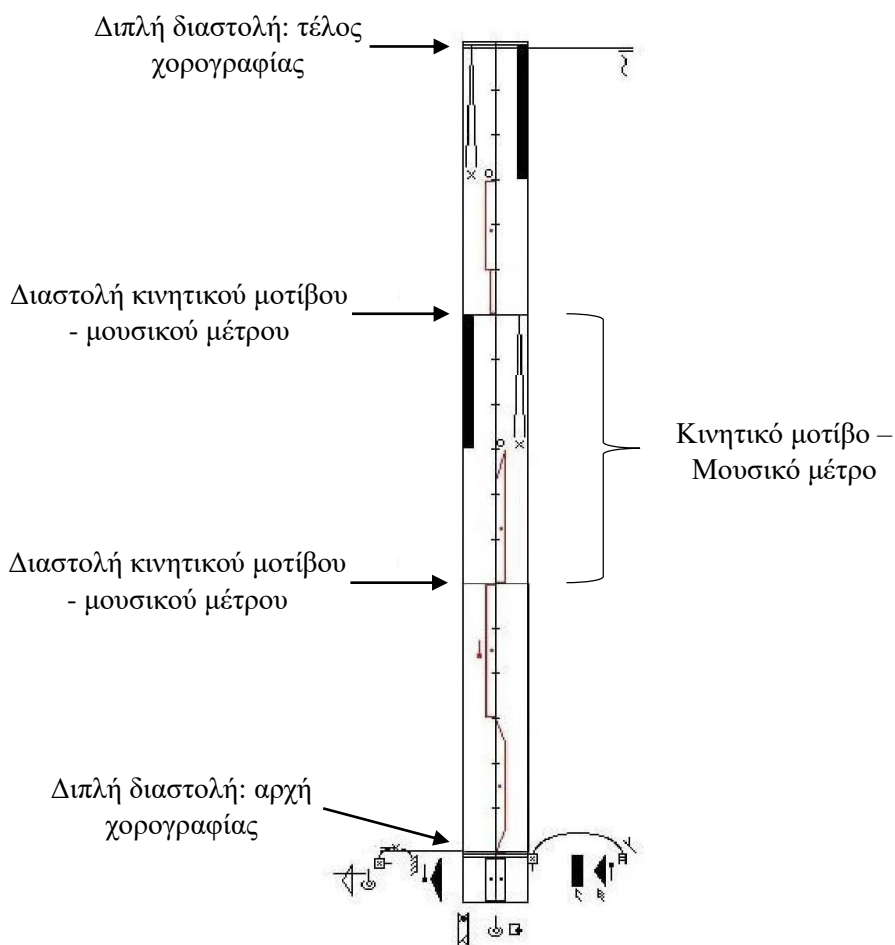
επανάληψη. Παρόλα αυτά, δεν κρίνεται σκόπιμη η περαιτέρω ανάλυσή τους στα πλαίσια της εν λόγω εργασίας.

Το Labanotation γράφεται και διαβάζεται από κάτω προς τα πάνω. Επομένως, οι κινήσεις που είναι γραμμένες στον κάθετο άξονα εκτελούνται διαδοχικά, η μία μετά την άλλη, ενώ αυτές που είναι γραμμένες στον οριζόντιο άξονα, εκτελούνται ταυτόχρονα.



Σχήμα 2.6: Αρχές Labanotation: α) διαδοχικές κινήσεις, β) ταυτόχρονες κινήσεις. (Πηγή [44])

Τέλος, κάτι το οποίο είναι σημαντικό να αναφερθεί σχετίζεται με την αποτύπωση του ρυθμού και της μουσικής φράσης σε ένα κινησιόγραμμα. Πιο συγκεκριμένα, η κεντρική γραμμή του, εκτός από τον κεντρικό άξονα του σώματος συνιστά, επίσης, και τη γραμμή της ροής του χρόνου, η οποία διαιρείται σε μικρά μέρη με παύλες, καθένα από τα οποία δηλώνει τη διάρκεια του μουσικού χτύπου, αλλά και τη διάρκεια του κινητικού στοιχείου. Η διαίρεση αυτή είναι συμβατική και ορίζεται σε κάθε περίπτωση από το/τη σημειογράφο [4]. Εκτός από αυτές τις παύλες, σε μια παρτιτούρα χορού, υπάρχουν και οριζόντιες γραμμές που ξεκινούν από την πρώτη και καταλήγουν στην τρίτη στήλη του τρίγραμμου στύλου, οι οποίες ονομάζονται διαστολές κινητικού μοτίβου και ταυτίζονται με τις διαστολές του μουσικού μέτρου που έχει οριστεί για τον εκάστοτε χορό. Έτσι, η μονή οριζόντια γραμμή αντιστοιχεί στα ρυθμικά μοτίβα και οι παύλες τα χωρίζουν σε επιμέρους κομμάτια, προσδιορίζοντας τους χτύπους της μουσικής. Ωστόσο, να σημειωθεί ότι υπάρχουν και διπλές οριζόντιες γραμμές, οι οποίες συνήθως εντοπίζονται στην αρχή και το τέλος ενός κινησιογράμματος, καθορίζοντας αντιστοίχως την αρχή και το τέλος της χορευτικής και της μουσικής φράσης ή ενός μείζονος τμήματος αυτής. Στην παρακάτω εικόνα επισημαίνουμε τα στοιχεία που προαναφέρθηκαν, παρουσιάζοντας ένα κινησιόγραμμα χορού, το οποίο προέρχεται από το σύνολο δεδομένων που χρησιμοποιείται στα πλαίσια της διπλωματικής εργασίας.



Σχήμα 2.7: Κινησιόγραμμα του χορού Ζωναράδικος.

## 2.4 Χορός και Τεχνολογία

Πρόκειται για δύο κόσμους διαφορετικούς που μέχρι σχετικά πρόσφατα δεν είχαν κάποιου είδους αλληλεπίδραση. Πράγματι, μέχρι τον 20<sup>ο</sup> αιώνα οι τεχνολογικές εξελίξεις δεν είχαν σημαντική επιρροή στον χώρο του χορού, πέραν, ίσως, της θεματολογίας ορισμένων χορογραφιών που παρουσιάζονταν σε ευρύ κοινό. Στις αρχές, ωστόσο, του 20<sup>ου</sup> αιώνα, του αιώνα των μεγάλων τεχνολογικών ανακαλύψεων, ορισμένοι καλλιτέχνες και επιστήμονες εμπνεύστηκαν από την ανάπτυξη της τεχνολογίας και χρησιμοποίησαν δημιουργικά τα διάφορα επιτεύγματα. Ανακάλυψαν εναλλακτικούς τρόπους ανάδειξης και εκμετάλλευσης των δυνατοτήτων της σε σχέση με την τέχνη και την ανθρώπινη έκφραση [48].

Ιστορικά, μία από τις ιέρειες του μοντέρνου χορού, η Loie Fuller που έδρασε στις αρχές του 1900, είναι εκείνη που πρώτη χρησιμοποιεί την τεχνολογία της εποχής, τις λάμπες

πυρακτώσεως, δηλαδή τον ηλεκτρισμό, για να τραβήξει το ενδιαφέρον του κοινού. Αργότερα, κατά τη δεκαετία του 1950, ο Alwin Nikolais, επίσης πειραματίζεται με διάφορα εργαλεία τεχνολογικής φύσης, χρησιμοποιώντας βιντεοπροβολές, οπτικά εφέ, και περίπλοκους φωτισμούς στις παραστάσεις που ανεβάζει [49]. Κατά το 1960 – 1970, η Trisha Brown ανεβάζει την παράσταση *Homemade*, η οποία μένει στην ιστορία. Κατά τη διάρκειά της εκτελούσε καθημερινές εργασίες, ενώ ταυτοχρόνως ένας προβολέας ταινιών, που ήταν τυλιγμένος στην πλάτη της, πρόβαλλε σε μια λευκή οθόνη πίσω από την πλάτη της, αλλά και σε όλους τους τοίχους της σκηνής, μια ταινία του Robert Whitman, στην οποία φαινόταν η ίδια να χορεύει την ίδια χορογραφία [50].



Σχήμα 2.8: Στιγμιότυπο από την παράσταση *Homemade* της Trisha Brown (1966). (Πηγή [51])

Προς τα τέλη του 20ου αιώνα, εμφανίζεται ο Merce Cunningham, σημαντικός χορευτής και χορογράφος της εποχής, πρωτοπόρος και εφευρέτης που προωθεί τη χρήση μαγνητικής ταινίας για τις συνθέσεις του, η οποία μέσω της επεξεργασίας του αναλογικού ηχητικού σήματος παρήγαγε ήχο. Στην πορεία, άρχισε να χρησιμοποιεί και μέσα δημιουργίας ηλεκτρονικής μουσικής, η οποία σταδιακά έγινε η μουσική συνοδεία στις χορογραφίες του [52]. Επιπλέον, χρησιμοποίησε ηλεκτρονικό υπολογιστή και ειδικό λογισμικό για τη δημιουργία κινητικών φράσεων, αλλά και για το διάνοιγμα νέων χορογραφικών δυνατοτήτων ευρύτερα. Δημιούργησε, μάλιστα, ένα εξαιρετικά χρήσιμο πρόγραμμα, το οποίο ονομάστηκε *Lifeforms*, και επέτρεπε την εικονική αναπαράσταση χορευτών στον υπολογιστή, διευκολύνοντας τόσο το χορογράφο, εφόσον μέσω αυτού μπορούσε να δοκιμάσει χορευτικές κινήσεις και διαφορετικούς συνδυασμούς για να οργανώσει τη χορογραφία, όσο και τους χορευτές, καθώς δε χρειαζόταν να εκτελεστεί

πολλές φορές ο εκάστοτε χορευτικός συνδυασμός. Με αυτή τη χορογραφική μέθοδο, ήταν σε θέση να δημιουργεί και να συνδυάζει διαφορετικές κινήσεις και φιγούρες απεριόριστα, χωρίς να χρειάζεται να αντιμετωπίσει τους περιορισμούς του τι είναι φυσιολογικό ή φυσικότερο για το σώμα των χορευτών [53]. Εκτός του *Lifeforms*, ο Cunningham ανέπτυξε το *Motion Capture*, το οποίο μετρούσε τη θέση και τον προσανατολισμό των χορευτών στον τρισδιάστατο χώρο σε πραγματικό χρόνο, περνώντας συγχρόνως τα δεδομένα αυτά στον υπολογιστή [54]. Οι χορευτές είχαν αισθητήρες σε διάφορα σημεία του σώματός τους και στις αρθρώσεις τους, οι οποίοι μετέφεραν τα σήματα στον υπολογιστή και εμφανίζονταν σαν κουκίδες. Σαν τελικό αποτέλεσμα, αυτές οι φιγούρες από κουκίδες έμοιαζαν με σκελετό ανθρωποειδούς ή αρχαϊκό είδωλο χωρίς απτά σημεία αρθρώσεων, οι οποίες μετατρέπονταν, μέσω ψηφιακών 3D προγραμμάτων, σε φιγούρες που έμοιαζαν να έχουν σχεδιαστεί με το χέρι [52].



Σχήμα 2.9: Χορεύτριες και εικονικές φιγούρες στην παράσταση *Biped* του Merce Cunningham (1999).  
(Πηγή [55])

Με την πάροδο του χρόνου, και καθώς η τεχνολογία εξελίσσεται με γρήγορους ρυθμούς, τείνει να εμπλέκεται ακόμη περισσότερο στον κόσμο του χορού, εμπλουτίζοντας τις χορογραφίες και προσφέροντας εντυπωσιακά θεάματα στο κοινό. Πλέον πολλοί χορογράφοι επιλέγουν να ανεβάσουν διαδραστικές παραστάσεις, όπου οι χορευτές «συνδιαλέγονται» με την τεχνολογία. Άλλοι χορογράφοι, ανάμεσα στους οποίους και η Siobhan Davies, ο Emio Greco και η Deborah Hay, συνεργάζονται με ερευνητικές ομάδες

σε προγράμματα, όπως το Siobhan Davies Replay, το Inside Movement Knowledge και το Motion Bank για τη χρήση τεχνολογικών μέσων που θα διευκολύνουν τη μελέτη σε σχέση με την καταγραφή της κίνησης και του χορού [5].

Κάτι που χαρακτηρίζει, ωστόσο, την εποχή μας σε μεγάλο βαθμό είναι οι αλματώδεις εξελίξεις στο πεδίο της Τεχνητής Νοημοσύνης και της Μηχανικής Μάθησης, οι οποίες σε συνδυασμό με την δυνατότητα επεξεργασίας του χορού σε ψηφιακή πλέον μορφή, γεννούν νέες προκλήσεις στον χώρο. Μερικές από τις πιο ευρέως διαδεδομένες εφαρμογές της Μηχανικής Μάθησης στον κόσμο του χορού είναι οι εξής:

- **Χορογραφία:** οι αλγόριθμοι αυτοί μπορούν να χρησιμοποιηθούν σαν «βοηθός» χορογράφου, προτείνοντας χορευτικούς συνδυασμούς που βασίζονται στο στυλ, στο οποίο έχουν εκπαιδευτεί κάθε φορά.
- **Διαδραστικές παραστάσεις Χορού:** οι αλγόριθμοι Μηχανικής Μάθησης μπορούν να εκπαιδευτούν στην αναγνώριση χορευτικών κινήσεων, κάτι το οποίο μπορεί να χρησιμοποιηθεί σε διαδραστικές παραστάσεις χορού για τον έλεγχο της μουσικής ή ακόμη και των οπτικών εφέ.
- **Εκπαίδευση στο χορό:** έχουν δημιουργηθεί προγράμματα που μπορούν να χρησιμοποιηθεί για εκπαιδευτικούς σκοπούς στα πλαίσια του χορού, τα οποία αναγνωρίζουν τη στάση, τη φόρμα και τις κινήσεις των μαθητών, τα αναλύουν και, εν τέλει, παρέχουν προσαρμοσμένες ασκήσεις, ιδέες ή ακόμη και διορθώσεις στους μαθητές. Έτσι, μπορεί να διευκολυνθεί η διδασκαλία του χορού, συμβάλλοντας στην καλύτερη κατανόηση της τεχνικής και στην αποτελεσματική εξάσκηση των ενδιαφερόμενων.
- **Ταξινόμηση ειδών χορού:** αξιοποιώντας τις διαφορετικές κινήσεις και τα ιδιαίτερα χαρακτηριστικά που παρατηρούνται σε κάθε χορευτικό είδος, οι αλγόριθμοι αυτοί μπορούν να καταλήξουν σε μια αξιόπιστη ταξινόμησή τους, μέσα από εικόνες, βίντεο ή ακόμη και παραστάσεις (live performances).
- **Σημειογραφία του χορού:** οι αλγόριθμοι Μηχανικής Μάθησης μπορούν να διευκολύνουν τη διαδικασία καταγραφής της χορευτικής κίνησης, αποκωδικοποιώντας τα «βήματα» του κάθε χορού και παράγοντας αυτόματα τα αντίστοιχα κινήσιογράμματα. Επιπλέον, μπορούν να χρησιμοποιηθούν για ταξινόμηση ή συσταδοποίηση διαφορετικών ειδών χορού, αναλύοντας τις αντίστοιχες παρτιτούρες χορού.

Με αυτήν την τελευταία εφαρμογή της Μηχανικής Μάθησης θα ασχοληθούμε στα πλαίσια της εν λόγω διπλωματικής εργασίας, αποσκοπώντας στη συσταδοποίηση κινησιογραμμάτων που προέρχονται από το είδος του Ελληνικού Παραδοσιακού Χορού. Για την επεξεργασία αυτών των πολύτιμων δεδομένων και την ανακάλυψη συσχετίσεων μεταξύ τους, θα χρησιμοποιηθούν τεχνικές Βαθιάς Μάθησης, οι έννοιες των οποίων αναλύονται στο επόμενο κεφάλαιο. Με αυτόν τον τρόπο, θα έρθουμε σε μια πρώτη επαφή με την υποκείμενη δομή τους, υπό την έννοια ότι κινησιογράμματα, στα οποία εντοπίζονται παρόμοια πρότυπα (patterns) και μοτίβα, ακόμη και μη ορατά μέσω απλής παρατήρησης, κατατάσσονται σε ίδιες ομάδες. Επιπροσθέτως, μέσω αυτής της συσταδοποίησης και της ερμηνείας των αποτελεσμάτων της, θα μας δοθεί η ευκαιρία διερεύνησης της συμπεριφοράς των χρησιμοποιούμενων τεχνικών σε ένα αρκετά πολύπλοκο σύνολο δεδομένων, λόγω της εξειδίκευσης που παρουσιάζει και των ομοιοτήτων μεταξύ των δειγμάτων του.

Συνολικά η μέθοδος που επιλέχθηκε να υλοποιηθεί στο πειραματικό μέρος αυτής της εργασίας, μπορεί να φανεί ιδιαίτερος βοηθητική στο κομμάτι της ανάλυσης δεδομένων (Data Analysis). Λαμβάνοντας, μάλιστα, υπόψη ότι πρόκειται για ένα πρωτότυπο πρόβλημα, το οποίο δεν έχει ακόμη βγει στο φως των δημοσιοποιημένων ερευνών, οποιεσδήποτε παρατηρήσεις και συμπεράσματα εξαχθούν, θα είναι ιδιαίτερος σημαντικά, ανοίγοντας το δρόμο για περαιτέρω εμβάθυνση.

Πράγματι, οι τεχνικές του επιστημονικού κλάδου της Τεχνητής Νοημοσύνης μπορούν να δώσουν νέες διαστάσεις στο κομμάτι της σημειογραφίας του χορού, παρά των δυσκολιών του ως αντικείμενο, με την εν λόγω έρευνα να αποτελεί μόνο την αρχή.

## Κεφάλαιο 3

### 3 Θεωρητικό Υπόβαθρο

---

Στο παρόν κεφάλαιο παρουσιάζονται οι θεωρητικές έννοιες και το επιστημονικό υπόβαθρο, τα οποία είναι απαραίτητα για την καλύτερη κατανόηση των τεχνικών που θα χρησιμοποιηθούν στο πρακτικό μέρος της εργασίας. Δίνονται οι ορισμοί και αναλύονται οι βασικές αρχές που διέπουν τη λειτουργία των συστημάτων της Τεχνητής Νοημοσύνης, της Μηχανικής Μάθησης, της Βαθιάς Μάθησης και των Νευρωνικών Δικτύων, ενώ ακόμη εξηγείται η εφαρμογή των μεθόδων της Συσταδοποίησης, της Μείωσης Διαστατικότητας και της Μεταφοράς Μάθησης.

#### 3.1 Τεχνητή Νοημοσύνη

Τεχνητή Νοημοσύνη – TN (Artificial Intelligence - AI) είναι ο κλάδος της Επιστήμης των Υπολογιστών, ο οποίος εξετάζει τη δημιουργία υπολογιστικών προγραμμάτων και εν γένει λογισμικού που μπορεί να πραγματοποιήσει έξυπνες ενέργειες [56]. Πραγματεύεται τη δημιουργία μηχανών που λειτουργούν με αυτόνομο τρόπο σε πολύπλοκα και διαρκώς μεταβαλλόμενα περιβάλλοντα [57]. Με άλλα λόγια, η TN θα μπορούσε να οριστεί ως η ικανότητα ενός συστήματος να ερμηνεύει δεδομένα που του παρέχονται, να εκπαιδεύεται και να μαθαίνει μέσω αυτών, έτσι ώστε, τελικά, να χρησιμοποιήσει τη γνώση που κατέκτησε για να επιτύχει συγκεκριμένες εργασίες – στόχους [58]. Ιστορικά, έχουν υπάρξει διαφορετικοί ορισμοί για την TN, οι οποίοι θα μπορούσαν να ομαδοποιηθούν, καταλήγοντας άλλοτε σε ανθρωποκεντρικές και άλλοτε σε ορθολογικές προσεγγίσεις για την επιστήμη αυτή [57].

Πρώτος, ο οποίος αναφέρθηκε σε νοήμονα συμπεριφορά υπολογιστών ήταν ο Alan Turing που μέσω της ομώνυμης δοκιμασίας (δοκιμασία *Turing*) έθεσε τη βάση για την αναγνώριση τέτοιου είδους νοημοσύνης σε τεχνητά συστήματα. Ωστόσο, η καθιέρωσή της ως επιστημονικό πεδίο έγινε το 1956 από τους John McCarthy και Marvin Minsky που διοργάνωσαν μια δίμηνη συνάντηση εργασίας στο Dartmouth College [59]. Έκτοτε, έχει γίνει αντικείμενο μελέτης πολλών ερευνητών κι έχει εξελιχθεί σε ένα εργαλείο χρήσιμο όχι

μόνο για τον κόσμο της επιστήμης, αλλά και για τη διευκόλυνση των ανθρώπων στην καθημερινότητά τους. Η Τεχνητή Νοημοσύνη βρίσκει πλέον εφαρμογή και επηρεάζει πολλούς τομείς της ζωής μας, ενώ αναμένεται να επιφέρει ακόμη μεγαλύτερες αλλαγές τα επόμενα χρόνια.

Πιο συγκεκριμένα, αναφορικά με τις εφαρμογές της, ανάμεσα σε αυτές ξεχωρίζουν η Μηχανική Μάθηση (Machine Learning), η οποία θα αναλυθεί εκτενώς στο επόμενο κεφάλαιο και η Όραση Υπολογιστών (Computer Vision), με αξιοσημείωτα αποτελέσματα στους τομείς της Υγείας, της Οικονομίας, της Ρομποτικής, της Βιομηχανίας, καθώς και πολλούς άλλους τομείς της καθημερινής ζωής. Βέβαια, η εκτεταμένη χρήση της έχει εγείρει αρκετές αντιδράσεις και αμφιβολίες, οδηγώντας τους ειδικούς να μιλούν για ηθικές και κοινωνικές επιδράσεις της ΤΝ στη ζωή των ανθρώπων. Σαφώς, έχει επιφέρει αλλαγές στην αγορά εργασίας κι έχει θέσει ηθικά διλήμματα, όπως αυτά περί ασφάλειας και ιδιωτικότητας. Ωστόσο, τα πλεονεκτήματα που προσφέρει είναι πολυάριθμα και ζητούμενο είναι η σωστή και ηθική αξιοποίησή τους.

### 3.2 Μηχανική Μάθηση

Μηχανική Μάθηση (Machine Learning - ML) είναι ένας κλάδος της Τεχνητής Νοημοσύνης, ο οποίος ασχολείται με τη μελέτη αλγορίθμων που στόχο έχουν να βελτιώσουν της επίδοσή τους σε προβλήματα που τους ανατίθενται, καθώς και να κάνουν ακριβείς προβλέψεις, χρησιμοποιώντας την εμπειρία που έχουν ήδη αποκτήσει. Η εμπειρία αυτή αφορά πληροφορίες που έχουν συλλεχθεί και διατίθενται για ανάλυση υπό τη μορφή ηλεκτρονικών δεδομένων. Όσο μεγαλύτερη σε μέγεθος και καλύτερης ποιότητας είναι αυτή η προϋπάρχουσα «εμπειρία», τόσο πιο επιτυχημένες θα προκύπτουν και οι προβλέψεις [60].

Το επιστημονικό αυτό πεδίο, λοιπόν, εξετάζει τη δημιουργία υπολογιστικών συστημάτων ικανών να μαθαίνουν, να σχηματίζουν μοντέλα – πρότυπα από ένα σύνολο δεδομένων και έτσι να βελτιώνουν την απόδοσή τους. Ακολουθούν τη διαδικασία της Επαγωγικής Μάθησης (Inductive Learning). Σύμφωνα με αυτή, ο άνθρωπος στην προσπάθειά του να κατανοήσει το περιβάλλον γύρω του και να μάθει, εν τέλει, από αυτό, δημιουργεί νοητικά μοντέλα (mental models), δηλαδή απλοποιημένες – αφαιρετικές εκδοχές του, ενώ στη συνέχεια συνδυάζοντας και οργανώνοντας τις παρατηρήσεις του αυτές, φτιάχνει νέες δομές, τα νοητικά πρότυπα (mental patterns) [61]. Με αυτόν τον τρόπο, έχουν σχεδιαστεί και οι αλγόριθμοι της Μηχανικής Μάθησης, έτσι ώστε να



αξιοποιούν τη γνώση που έχουν λάβει, να δημιουργούν συσχετίσεις μεταξύ παρατηρήσεων και αποτελεσμάτων και, τελικά, να χρησιμοποιούν τα πρότυπα για να κάνουν σωστές προβλέψεις.

Στον όρο αυτό πρώτος αναφέρθηκε ο Arthur Samuel το 1959, μέσα από ένα παιχνίδι (game of Checkers), δίνοντας τις βασικές αρχές της Μηχανικής Μάθησης, οι οποίες θα μπορούσαν, σύμφωνα με εκείνον, να χρησιμοποιηθούν και σε άλλα προβλήματα [62]. Αργότερα, έναν ευρέως χρησιμοποιούμενο ορισμό έδωσε ο Tom Mitchell: «Ένα πρόγραμμα υπολογιστή θεωρείται ότι μαθαίνει από την εμπειρία  $E$  σε σχέση με μια κατηγορία εργασιών  $T$  και μια μετρική απόδοσης  $P$ , αν η απόδοση του σε εργασίες της  $T$ , όπως μετριοούνται από την  $P$ , βελτιώνονται με την εμπειρία  $E$ » [63].

Η Μηχανική Μάθηση, στις μέρες μας, ωστόσο, δεν έχει σχέση με τη Μηχανική Μάθηση που προσπάθησαν να περιγράψουν οι άνθρωποι αυτοί πριν κάποια χρόνια. Φυσικά, η βάση είναι η ίδια, όμως η αλματώδης ταχύτητα με την οποία εξελίσσεται, την έχει ανάγκη σε εργαλείο που χρησιμοποιείται από μια πληθώρα υπολογιστικών εργασιών και, πλέον, αποτελεί αναπόσπαστο κομμάτι της καθημερινής ζωής των ανθρώπων. Με τους αλγορίθμους της, έχει καταστεί δυνατή η γρήγορη και αυτόματη παραγωγή μοντέλων για την ανάλυση πολύ μεγάλων και εξαιρετικά σύνθετων συνόλων δεδομένων, ενώ παράλληλα παρέχονται αξιόπιστα αποτελέσματα ακόμη και σε περίπλοκα προβλήματα.

Μερικά χαρακτηριστικά παραδείγματα εφαρμογών της είναι: η Επεξεργασία Φυσικής Γλώσσας (Natural Language Processing), η Οπτική Αναγνώριση Χαρακτήρων (Optical Character Recognition), τα Συστήματα Συστάσεων (Recommendation Systems), η Διάγνωση Ασθενειών, διάφορα προβλήματα Όρασης Υπολογιστών, η Αναγνώριση Φωνής και Εικόνας (Speech / Image Recognition), η Ανίχνευση Απάτης (Fraud Detection), τα Αυτόνομα Οχήματα (Autonomous Vehicles), καθώς και η Ομαδοποίηση Εικόνων (Image Clustering), με το οποίο και θα ασχοληθούμε στα πλαίσια της παρούσας εργασίας.

Τέλος, να αναφερθεί ότι η Μηχανική Μάθηση χωρίζεται σε 3 βασικές κατηγορίες, ανάλογα με το είδος των δεδομένων και τη μέθοδο που χρησιμοποιείται κατά την εκπαίδευση του συστήματος, την Επιβλεπόμενη, τη Μη Επιβλεπόμενη και την Ενισχυτική Μάθηση, οι οποίες αναλύονται παρακάτω.

### 3.2.1 Επιβλεπόμενη Μάθηση

Κατά την Επιβλεπόμενη Μάθηση (Supervised Learning), στο σύστημα παρέχονται επισημασμένα δεδομένα (labeled data), βάσει των οποίων εκπαιδεύεται και, στη συνέχεια,

κάνει προβλέψεις για νέα δεδομένα, τα οποία δεν έχει συναντήσει ξανά. Πιο συγκεκριμένα, ως είσοδος δίνεται ένα σύνολο δεδομένων εκπαίδευσης, το οποίο αποτελείται από ένα σύνολο μεταβλητών  $x$  και ένα σύνολο ετικετών  $y$  που χαρακτηρίζουν κάθε δείγμα εισόδου. Στόχος είναι να δημιουργηθεί μια συνάρτηση  $y = f(x)$  (mapping function), τέτοια ώστε ο αλγόριθμος να προβλέπει σωστά σε μη επισημασμένη είσοδο, βασιζόμενος στις πληροφορίες που έχει συλλέξει από την φάση της εκπαίδευσης, δηλαδή να έχει τη δυνατότητα γενίκευσης (generalisation). Τα προβλήματα που προσεγγίζονται με αλγορίθμους Επιβλεπόμενης Μάθησης συνήθως αφορούν 2 βασικές υποκατηγορίες της, την Ταξινόμηση και την Παλινδρόμηση.

### 3.2.2 Μη Επιβλεπόμενη Μάθηση

Κατά τη Μη Επιβλεπόμενη Μάθηση (Unsupervised Learning), στο σύστημα παρέχονται μη επισημασμένα δεδομένα (unlabeled data), στα οποία βασίζεται ώστε να ανακαλύψει κρυμμένα μοτίβα, τα οποία θα το βοηθήσουν να κάνει σωστές προβλέψεις. Σε αυτήν την περίπτωση, δηλαδή, η μηχανή εκπαιδεύεται σε πρότυπα εισόδου που δεν έχουν συγκεκριμένες τιμές εξόδου [57]. Συχνά, παρομοιάζεται με τη διαδικασία μάθησης που συμβαίνει στον ανθρώπινο εγκέφαλο, καθώς οι άνθρωποι, επίσης, μαθαίνουν, παρατηρώντας τον κόσμο γύρω τους και βρίσκοντας μοτίβα, τα οποία προσπαθούν να κατανοήσουν ανακαλύπτοντας σχέσεις και σχηματίζοντας μοντέλα – περιγραφές των αντικειμένων [64]. Στόχος της μη Επιβλεπόμενης Μάθησης είναι η εύρεση της υποκείμενης δομής του συνόλου δεδομένων, η αναγνώριση εκείνων των προτύπων που θα διευκολύνουν τη δημιουργία συσχετίσεων ανάμεσά τους, ώστε να γίνει εφικτή η ομαδοποίηση βάσει των ομοιοτήτων και των διαφορών τους. Εν τέλει, οι αλγόριθμοι αυτού του είδους Μηχανικής Μάθησης κατασκευάζουν μοντέλα για κάποιο σύνολο δεδομένων υπό μορφή παρατηρήσεων, χωρίς να υπάρχει γνώση για τις επιθυμητές εξόδους [61]. Τα προβλήματα που προσεγγίζονται με αλγορίθμους μη Επιβλεπόμενης Μάθησης συνήθως αφορούν 3 βασικές υποκατηγορίες της, τη Συσταδοποίηση, την Ανάλυση Συσχετισμών και τη Μείωση Διαστατικότητας.

#### Συσταδοποίηση (Clustering)

Σε αυτήν την υποκατηγορία ανήκουν τα προβλήματα ομαδοποίησης των δεδομένων σε μικρότερα ομοιογενή υποσύνολα [60]. Τα δείγματα που μοιάζουν μεταξύ τους ταξινομούνται στην ίδια συστάδα, ενώ εκείνα με χαρακτηριστικά ανόμοια ταξινομούνται

σε διαφορετικές συστάδες [65]. Σκοπός των αλγορίθμων συσταδοποίησης είναι η εξαγωγή χαρακτηριστικών, η αναγνώριση προτύπων και ο εντοπισμός των μοτίβων εκείνων που διευκολύνουν το δίκτυο να κάνει το διαχωρισμό των δειγμάτων σε συστάδες (clusters). Χαρακτηριστικό παράδειγμα αποτελούν τα Συστήματα Συστάσεων (Recommendation Systems), τα οποία βρίσκουν εφαρμογή σε πολλούς τομείς της σύγχρονης ζωής και αποσκοπούν στην εύρεση των προτιμήσεων του χρήστη, ώστε να του κάνουν προτάσεις που περισσότερο τον ενδιαφέρουν.

### Ανάλυση Συσχετισμών (Association Rules)

Σε αυτήν την υποκατηγορία ανήκουν τα προβλήματα εύρεσης συσχετισμών και εξαρτήσεων μεταξύ των δειγμάτων ενός μεγάλου συνόλου δεδομένων [66]. Η ανακάλυψη κανόνων συσχέτισης παρέχει σημαντικές πληροφορίες για τα εκάστοτε δεδομένα και γι' αυτό το λόγο η Ανάλυση Συσχετισμών συχνά χρησιμοποιείται από τις επιχειρήσεις για τη λήψη αποφάσεων. Η μέθοδος αυτή έκανε την εμφάνισή της αρκετά αργότερα από τη Μηχανική Μάθηση, ενώ έχει επηρεαστεί σημαντικά από τον ερευνητικό κλάδο των βάσεων δεδομένων [67]. Ένα χαρακτηριστικό παράδειγμα είναι η Ανάλυση Καλαθιού Αγοράς (Market Basket Analysis), όπου γίνεται μελέτη για την εύρεση μοτίβων σχετικά με τις αγορές των πελατών, βρίσκοντας συσχετίσεις μεταξύ των αντικειμένων που τοποθετούν οι πελάτες στα καλάθια τους.

### Μείωση Διαστατικότητας (Dimensionality Reduction)

Σε αυτήν την υποκατηγορία ανήκουν τα προβλήματα μείωσης των διαστάσεων ενός συνόλου δεδομένων, με ταυτόχρονη διατήρηση των βασικών του χαρακτηριστικών και κατ' επέκταση της ακεραιότητάς του. Στόχος είναι να ελαττωθεί η πολυπλοκότητα αυτών των δεδομένων, για να καταστεί ευκολότερη η επεξεργασία, η ανάλυση, η κατανόηση και η οπτικοποίησή τους και, φυσικά, να μειωθεί η απαίτηση σε υπολογιστικούς πόρους. Χρησιμοποιείται συχνά στα στάδια προεπεξεργασίας των δεδομένων για τη βελτίωση της απόδοσης των αλγορίθμων μηχανικής μάθησης, ενώ χαρακτηριστικά παραδείγματα εφαρμογής της αποτελούν η Επιλογή Χαρακτηριστικών (Feature Selection) και η Οπτικοποίηση Δεδομένων (Data Visualisation).

Η Μείωση Διαστατικότητας, καθώς και η Συσταδοποίηση, θα αναλυθούν εκτενέστερα σε επόμενες υποενότητες, καθώς αποτελούν σημαντικά σημεία της παρούσας διπλωματικής εργασίας.

### 3.2.3 Ενισχυτική Μάθηση

Κατά την Ενισχυτική Μάθηση (Reinforcement Learning), το σύστημα εκπαιδεύεται συγκεντρώνοντας πληροφορίες μέσω της αλληλεπίδρασης με το περιβάλλον του (environment). Παρατηρεί τον τρόπο λειτουργίας του, καταστρώνει μια σειρά ενεργειών και μαθαίνει βλέποντας τα αποτελέσματά τους, χωρίς να καθοδηγείται από κάποιον εξωτερικό επιβλέποντα. Η τεχνική αυτή έχει επηρεαστεί από μοντέλα που βασίζονται στη θεωρία της επιβράβευσης και τιμωρίας (reward and punishment) και λειτουργεί ως εξής: ο μαθητής – σύστημα βρίσκεται μέσα σε ένα χώρο κι έχει ένα συγκεκριμένο στόχο που δίνεται από το εκάστοτε πρόβλημα. Για να πετύχει τον στόχο αυτό εκτελεί μια σειρά κινήσεων. Σαν απάντηση σε κάθε ενέργεια λαμβάνει είτε μια ανταμοιβή, εάν η κίνησή του ήταν ωφέλιμη ως προς τον τελικό στόχο, είτε μια ποινή, εάν τον απομάκρυνε από αυτόν. Τελικός σκοπός της μηχανής είναι να μεγιστοποιήσει την ανταμοιβή του, συνυπολογίζοντας επιβραβεύσεις και τιμωρίες, καθορίζοντας έτσι τη βέλτιστη πορεία των κινήσεών του. Ωστόσο, πολλές φορές εμπίπτει σε δίλημμα και καλείται να διαλέξει ανάμεσα στην εξερεύνηση του περιβάλλοντος, δοκιμάζοντας άγνωστες ενέργειες για να αποκτήσει νέες πληροφορίες και την εκμετάλλευση των ήδη γνωστών πληροφοριών, ακολουθώντας «πεπατημένα μονοπάτια» που γνωρίζει ότι προσφέρουν επιβράβευση (exploration vs exploitation) [60].

Πρόκειται, λοιπόν, για μια τεχνική Μηχανικής Μάθησης που βασίζεται στην ανατροφοδότηση (feedback), όπου το σύστημα θα πρέπει μόνο του να ανακαλύψει τη σωστή στρατηγική ενεργειών, σε αντίθεση με τις παραπάνω περιπτώσεις, της Επιβλεπόμενης και Μη Επιβλεπόμενης, όπου βασιζόταν σε δεδομένα που του παρέχονταν. Συνήθως εφαρμογή βρίσκει σε προβλήματα σχεδιασμού (planning), σε διάφορους τομείς, όπως στην Ρομποτική, τον Αυτόματο Έλεγχο, τη Βιολογία, την Ιατρική, ακόμη και στην ανάπτυξη Παιχνιδιών για υπολογιστές. Χαρακτηριστικά παραδείγματα, όπου χρησιμοποιείται η Ενισχυτική Μάθηση είναι ο έλεγχος κίνησης Ρομπότ και η βελτιστοποίηση εργασιών σε εργοστασιακούς χώρους [67].

### 3.3 Βαθιά Μηχανική Μάθηση

Η Βαθιά Μηχανική Μάθηση (Deep Learning) αποτελεί υποσύνολο της Μηχανικής Μάθησης, οπότε και ασχολείται με προβλήματα παρόμοια με την τελευταία, έχοντας σαν στόχο την αποτελεσματικότερη εκπαίδευση των μηχανών, κάτι που έχει σαν συνέπεια, την αποδοτικότερη επίλυση των εργασιών που τους ανατίθενται. Έχει εμπνευστεί από τη δομή και τον τρόπο λειτουργίας των εγκεφαλικών κυττάρων του ανθρώπου, που ονομάζονται

νευρώνες. Όπως και ο ανθρώπινος εγκέφαλος, λοιπόν, έτσι και τα συστήματα Βαθιάς Μάθησης είναι δίκτυα πολλαπλών στρωμάτων, δηλαδή αποτελούνται από διαδοχικά «επίπεδα», καθένα από τα οποία περιλαμβάνει απλές μονάδες επεξεργασίας, και όλα μαζί συνθέτουν ένα λειτουργικό σύνολο που, εν τέλει, αποσκοπεί στην εξαγωγή χαρακτηριστικών (feature extraction) από τα δεδομένα που τους παρέχονται.

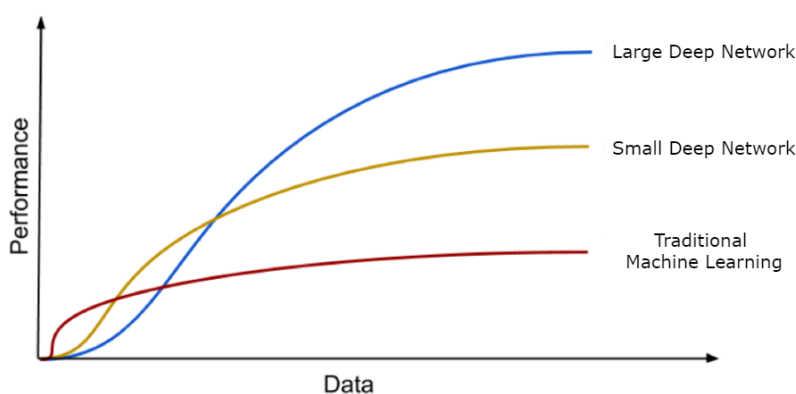


Σχήμα 3.1: Διάγραμμα συσχέτισης AI – ML – DL.

Τον όρο της Βαθιάς Μάθησης εισήγαγε πρώτος ο Geoffrey Hinton το 2006, και έκτοτε έχει γνωρίσει ραγδαία ανάπτυξη. Έχει βοηθήσει στην επίλυση προβλημάτων που προσπαθούσε καιρό να αποσαφηνίσει ο κλάδος της Τεχνητής Νοημοσύνης, ενώ αποδείχτηκε εξαιρετικά χρήσιμη στην ανακάλυψη περίπλοκων δομών σε υψηλών διαστάσεων δεδομένα [68]. Βρίσκει εφαρμογή σε πολλούς τομείς της σύγχρονης ζωής, ενώ αποδίδει πολύ υψηλά (state-of-the-art) αποτελέσματα σε εργασίες όπως, η αναγνώριση εικόνας και φωνής. Χαρακτηριστικά παραδείγματα, όπου χρησιμοποιούνται οι τεχνικές της Βαθιάς Μάθησης είναι τα αυτό-οδηγούμενα οχήματα, η αυτόματη μετάφραση, καθώς και η ανάλυση συναισθημάτων. Στην ευρεία διάδοσή της συνέβαλαν και βιβλιοθήκες Μηχανικής Μάθησης ανοικτού κώδικα (open source), όπως είναι η Tensorflow, που χρησιμοποιείται στα πλαίσια της παρούσας εργασίας, καθώς και η PyTorch, που παρέχουν έτοιμα υπολογιστικά components, διευκολύνοντας την επίλυση μεγάλου εύρους προβλημάτων.

Αυτό που διαφοροποιεί τα δύο είδη μάθησης, τη Βαθιά από τη Μηχανική, είναι ο τύπος των δεδομένων που χρησιμοποιούν, καθώς και η διαδικασία με την οποία μαθαίνουν. Αρχικά, στη Βαθιά Μάθηση, το σύστημα χρησιμοποιεί μη-επεξεργασμένα δεδομένα (raw

data), όπως εικόνες, κείμενο, φωνή, κτλ. και αυτόματα εξάγει τα χαρακτηριστικά, τις πληροφορίες εκείνες που θεωρεί σημαντικές για το εκάστοτε πρόβλημα. Αυτό έρχεται σε αντίθεση με τις συμβατικές τεχνικές Μηχανικής Μάθησης, όπου δεν μπορούν να γίνει απευθείας επεξεργασία πάνω σε τέτοιου είδους δεδομένα. Είναι απαραίτητη η συμβολή ενός ανθρώπου που θα έχει την τεχνογνωσία για να σχεδιάσει, σε πρώτη φάση, κάποιον αλγόριθμο που κάνει εξαγωγή χαρακτηριστικών και μετατρέπει τα raw data σε μια κατάλληλη εσωτερική αναπαράσταση [68], ώστε, σε δεύτερη φάση, να μπορούν να χρησιμοποιηθούν σαν είσοδος σε αλγόριθμους Μηχανικής Μάθησης. Ακόμη μία διαφορά τους, έγκειται στο ότι οι τεχνικές της Βαθιάς Μάθησης δύνανται να επεξεργαστούν μεγάλο όγκο δεδομένων, και, μάλιστα, όσα περισσότερα δεδομένα τους παρέχονται, τόσο αυξάνεται και η απόδοσή τους. Να σημειωθεί ότι τα πιο «βαθιά» δίκτυα, δηλαδή εκείνα που περιλαμβάνουν περισσότερα στρώματα τείνουν να εμφανίζουν βελτιωμένη συμπεριφορά σε σχέση με μικρότερα δίκτυα, όπως φαίνεται και στην παρακάτω εικόνα. Το πλεονέκτημά τους αυτό είναι που τις κάνει ιδιαίτερα δημοφιλείς τα τελευταία χρόνια, καθώς ζούμε στην εποχή των Μεγάλων Δεδομένων (Big Data). Αυτό, όμως, έρχεται σε αντιδιαστολή με τους παραδοσιακούς αλγορίθμους Μηχανικής Μάθησης που με την αύξηση της ποσότητας των δεδομένων φτάνουν σε ένα σημείο κορεσμού, από το οποίο και έπειτα δεν παρατηρείται βελτίωση των αποτελεσμάτων τους [69].



Σχήμα 3.2: Συγκριτική παρουσίαση της απόδοσης δικτύων Μηχανικής και Βαθιάς Μάθησης. (Πηγή [69])

### 3.4 Μείωση Διαστατικότητας

Είναι γεγονός ότι διανύουμε την «εποχή της πληροφορίας», μια περίοδο όπου υπάρχει άφθονη ροή δεδομένων από ποικίλες πηγές και σε διάφορες μορφές και η οποία χαρακτηρίζεται από μια έντονη ανταλλαγή πληροφοριών, ενώ η πρόσβαση και απόκτηση

γνώσεων σε πληθώρα τομέων είναι πιο εύκολη από ποτέ άλλοτε. Ο ολοένα αυξανόμενος αριθμός των σύγχρονων εφαρμογών παράγει μαζικά πλήθη δεδομένων. Τα δεδομένα αυτά που καλούμαστε να επεξεργαστούμε, ωστόσο, είναι στην πλειονότητά τους πολυδιάστατα [70]. Σαφώς, αυτός ο μεγάλος αριθμός διαστάσεων παρέχει μεγαλύτερο όγκο πληροφορίας, με αποτέλεσμα την καλύτερη κατανόηση των συσχετίσεων μεταξύ των δειγμάτων ενός συνόλου δεδομένων και κατ' επέκταση την υψηλότερη ακρίβεια, σε ορισμένες περιπτώσεις, των μοντέλων που τα χρησιμοποιούν. Από την άλλη πλευρά, όμως, η υψηλή διαστατικότητα σημαίνει και μεγάλη πολυπλοκότητα, κάτι που μπορεί να οδηγήσει σε προβλήματα όπως η υπερεκπαίδευση και η αδυναμία γενίκευσης, καθώς επίσης και σε υψηλές απαιτήσεις όσον αφορά τους υπολογιστικούς πόρους. Έτσι, συχνά, η απόδοση ορισμένων αλγορίθμων μειώνεται όσο αυξάνονται οι διαστάσεις των δεδομένων, ένα φαινόμενο που είναι γνωστό ως *κατάρα της διαστατικότητας* (*Curse of Dimensionality*).

Τα προβλήματα αυτά έχουν στρέψει το ενδιαφέρον της επιστημονικής κοινότητας στην ανάπτυξη μεθόδων Μείωσης της Διαστατικότητας, δηλαδή ελάττωσης των διαστάσεων των δεδομένων, με απώτερο σκοπό την πιο αποτελεσματική επεξεργασία τους. Αυτό επιτυγχάνεται μέσω της προβολής τους σε ένα μικρότερης διάστασης υποχώρο, ικανό να αποτυπώσει την «ουσία» [71], δηλαδή την απαραίτητη χρήσιμη πληροφορία. Με αυτόν τον τρόπο, παρέχεται μια πιο «συμπαγής» και εύκολα κατανοητή αναπαράσταση των δεδομένων, εστιάζοντας στις πιο σημαντικές μεταβλητές [72].

Οι τεχνικές αυτές βρίσκουν εφαρμογή σε πληθώρα τομέων, όπως είναι η προεπεξεργασία δεδομένων, η επιλογή χαρακτηριστικών, η αποθορυβοποίηση και η οπτικοποίηση δεδομένων, ενώ επίσης συμβάλλουν σημαντικά στην αύξηση της απόδοσης εργασιών όπως η Ταξινόμηση, η Συσταδοποίηση και η Επεξεργασία Φυσικής Γλώσσας.

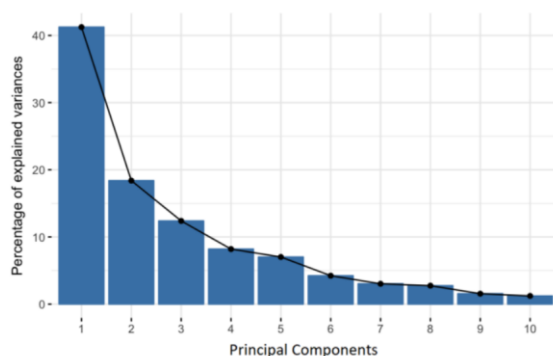
Στο παρόν κεφάλαιο, θα επικεντρωθούμε σε μια από τις πιο γνωστές μεθόδους του πεδίου αυτού, την Ανάλυση Κύριων Συνιστωσών (Principal Component Analysis), η οποία χρησιμοποιείται και στα πλαίσια της διπλωματικής εργασίας. Ανήκει στην κατηγορία των γραμμικών τεχνικών μείωσης διαστατικότητας και υλοποιεί τον μετασχηματισμό των αρχικών δειγμάτων σε ένα νέο χώρο μικρότερης διάστασης, όπου οι νέες μεταβλητές ονομάζονται κύριες συνιστώσες.

### **3.4.1 Ανάλυση Κύριων Συνιστωσών**

Η Ανάλυση Κύριων Συνιστωσών (Principal Component Analysis – PCA) είναι μια ευρέως διαδεδομένη τεχνική που επιτυγχάνει τη μείωση της διάστασης ενός συνόλου δεδομένων, διατηρώντας συγχρόνως μεγάλο όγκο της αρχικής πληροφορίας. Ανήκει στους

αλγορίθμους μη επιβλεπόμενης μάθησης, οπότε χρησιμοποιεί δεδομένα μη επισημασμένα και βρίσκει εφαρμογή σε προβλήματα μείωσης διαστατικότητας στη στατιστική αναγνώριση προτύπων, αλλά και την επεξεργασία σήματος.

Πιο αναλυτικά, η μέθοδος αυτή εκτελεί ορθογώνιο μετασχηματισμό ενός υψηλής διάστασης χώρου εισόδου, του οποίου τα στοιχεία είναι πιθανώς συσχετισμένα, σε έναν χαμηλής διάστασης χώρο, του οποίου τα στοιχεία είναι ασυσχέτιστα [73]. Οι ασυσχέτιστες αυτές μεταβλητές αποτελούν τις κύριες συνιστώσες (principal components), οι οποίες συμπυκνώνουν κατά φθίνοντα τρόπο τη μεταβλητότητα στα δεδομένα εισόδου που στην πραγματικότητα αντικατοπτρίζει το μέτρο του πληροφοριακού τους περιεχομένου. Έτσι, η πρώτη κύρια συνιστώσα (PC1) έχει τη μεγαλύτερη δυνατή διακύμανση, δηλαδή αντιπροσωπεύει το μεγαλύτερο μέρος της μεταβλητότητας των δεδομένων και κατ' επέκταση διατηρεί περισσότερες πληροφορίες από το αρχικό σύνολο δεδομένων, σε σχέση με τη δεύτερη κύρια συνιστώσα (PC2), η οποία έχει την αμέσως επόμενη μεγαλύτερη διακύμανση και δεν περιλαμβάνει τις πληροφορίες που εμπεριέχονται στην PC1. Κατά αυτόν τον τρόπο ορίζεται ο γραμμικός αυτός μετασχηματισμός, με κάθε επόμενη συνιστώσα PC3, PC4, κ.ο.κ. να περιέχει ολοένα και μικρότερο μέρος της πληροφορίας, να αντιστοιχίζεται στην υψηλότερη υπάρχουσα διακύμανση και, φυσικά, να είναι ορθογώνια (και επομένως ασυσχέτιστη) με τις προηγούμενες από αυτήν συνιστώσες.



Σχήμα 3.3: Ποσοστό περιεχόμενης πληροφορίας ανά κύρια συνιστώσα. (Πηγή [74])

Σε πρακτικό επίπεδο, η μέθοδος αυτή συχνά χρησιμοποιείται όταν το αρχικό σύνολο δεδομένων είναι μεγάλο, όπως συχνά συμβαίνει σε προβλήματα Μηχανικής Μάθησης, με στόχο αυτό να μπορέσει να αναπαρασταθεί μέσω ενός μειωμένου αριθμού «ωφέλιμων» χαρακτηριστικών. Η μέθοδος PCA, λοιπόν, επιτυγχάνει να μειώσει τη διαστατικότητα του αρχικού χώρου, προβάλλοντας τα δεδομένα στις κύριες συνιστώσες και έτσι συμπυκνώνει την πληροφορία. Με αυτόν τον τρόπο, παρέχεται όχι μόνο καλύτερη κατανόηση των πληροφοριών του συνόλου δεδομένων, αλλά παράλληλα, απομακρύνονται οι συνιστώσες

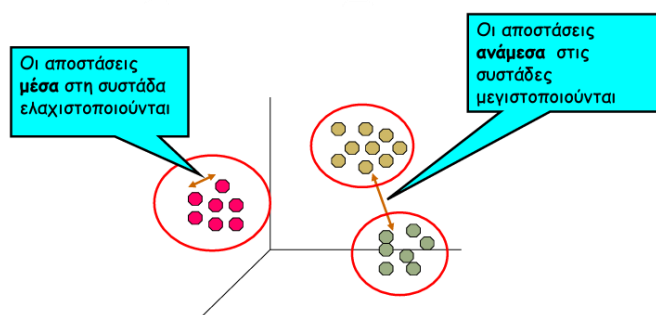


με χαμηλή διακύμανση (παραλείποντας πλεονάζουσες και λιγότερο σημαντικές πληροφορίες), μειώνοντας ακόμη και τον θόρυβο. Όλα αυτά συμβάλλουν στην αποτελεσματικότερη οπτικοποίηση των δεδομένων και συνεπώς σε πιο ολοκληρωμένη εικόνα όσον αφορά τη δομή τους και κατ'έκταση στην εξαγωγή συμπερασμάτων που θα διευκολύνουν τη μετέπειτα επεξεργασία τους.

Συνοψίζοντας, η PCA ως τεχνική προεπεξεργασίας, στο πλαίσιο της εξόρυξης δεδομένων, μπορεί να φανεί εξαιρετικά χρήσιμη, καθώς επιτυγχάνει τη δημιουργία ενός νέου συνόλου δεδομένων, μικρότερο του αρχικού, το οποίο, όμως, διατηρεί όσο το δυνατόν περισσότερη πληροφορία, αποδίδοντας τελικά (σχεδόν) τα ίδια αποτελέσματα με το σύνολο δεδομένων εισόδου. Με αυτόν τον τρόπο, τόσο η ταχύτητα όσο και η επίδοση αλγορίθμων Μηχανικής Μάθησης μπορούν να βελτιωθούν σημαντικά.

### 3.5 Συσταδοποίηση

Συσταδοποίηση (Clustering) είναι η διαδικασία διαμερισμού ενός συνόλου δεδομένων σε φυσικές ομαδοποιήσεις προτύπων, έτσι ώστε το μέτρο ομοιότητας μεταξύ οποιουδήποτε ζεύγους παρατηρήσεων αντιστοιχίζεται σε κάθε ομάδα, να ελαχιστοποιεί μια συγκεκριμένη συνάρτηση κόστους [75]. Με απλά λόγια, είναι μια μέθοδος, κατά την οποία το αρχικό σύνολο δειγμάτων χωρίζεται σε μικρότερες ομάδες, τις συστάδες (clusters), αναλόγως με το βαθμό ομοιότητας των δεδομένων μεταξύ τους. Στόχος είναι να δημιουργηθεί ένα σύνολο από ομάδες με υψηλή ομοιότητα εντός των ομάδων (intra-cluster similarity) και χαμηλή ομοιότητα μεταξύ των διαφορετικών ομάδων (inter-cluster similarity). Έτσι, η κάθε συστάδα θα απαρτίζεται από στοιχεία, καθένα από τα οποία θα είναι πιο «κοντά», δηλαδή θα μοιάζει περισσότερο, με κάθε άλλο στοιχείο της ίδιας συστάδας, σε σχέση με κάποιο άλλο στοιχείο διαφορετικής συστάδας [65].



Σχήμα 3.4: Intra-cluster και inter-cluster similarity. (Πηγή [76])

Πρόκειται για μία τεχνική μη Επιβλεπόμενης Μάθησης που συναντάται σε πολλά ακόμη επιστημονικά πεδία, συμπεριλαμβανομένων των επιστημών της ζωής (Βιολογία,

Ζωολογία) και της Γης (Γεωγραφία, Γεωλογία, Σεισμολογία), των ιατρικών (Ψυχιατρική, Παθολογία) και των κοινωνικών επιστημών (Κοινωνιολογία, Αρχαιολογία) [73], της Οικονομίας, της Αστροφυσικής και της Βιοπληροφορικής.

Εν γένει, η ομαδοποίηση αντικειμένων, προσώπων, εννοιών ή γεγονότων, αποτελεί μια από τις θεμελιακές διανοητικές δραστηριότητες του ανθρώπου, η οποία χρησιμοποιείται για την αποτελεσματική διαχείριση του μεγάλου όγκου πληροφοριών που δέχεται καθημερινά [73]. Μάλιστα, η ιδέα της ταξινόμησης παρόμοιων πραγμάτων σε κατηγορίες θεωρείται «πρωτόγονη» διεργασία, εφόσον ο προϊστορικός άνθρωπος συχνά ομαδοποιούσε αντικείμενα που μοιράζονταν ορισμένα χαρακτηριστικά [77]. Οπότε η συσταδοποίηση, ιστορικά, υπήρχε σαν πρόβλημα ήδη από την εποχή του Αριστοτέλη, ενώ εξετάστηκε εκτενέστερα κατά τον 18<sup>ο</sup> αιώνα από αναγνωρισμένους φυσιολόγους [78]. Η μελέτη αυτής της χρήσιμης ιδιότητας του εγκεφάλου του ανθρώπου σε επιστημονικό επίπεδο, ωστόσο, ξεκίνησε στις αρχές του 20<sup>ου</sup> αι. από τους κλάδους της Ανθρωπολογίας και της Ψυχολογίας για να διερευνηθεί αργότερα και στα πλαίσια του πεδίου της Επιστήμης των Υπολογιστών.

Αποτελεί, πλέον, μια ευρέως διαδεδομένη μέθοδο ανάλυσης δεδομένων, καθώς βοηθά στην καλύτερη κατανόηση της δομής τους, μέσω της εξαγωγής γνώσης, ανακαλύπτοντας πρότυπα και μοτίβα σε μη επισημασμένα δεδομένα. Δημοφιλέστερες εφαρμογές του clustering παρατηρούνται στον κλάδο της Μηχανικής Μάθησης και της Αναγνώρισης Προτύπων, με χαρακτηριστικά παραδείγματα τα Συστήματα Συστάσεων, την Ανάλυση Κοινωνικών Δικτύων (Social Network Analysis) και την Τμηματοποίηση της Αγοράς (Market Segmentation).

Να σημειωθεί ότι το clustering είναι μια διαδικασία «υποκειμενική», οπότε υπάρχουν αρκετοί τρόποι για να φτάσουμε σε κάποιο αποτέλεσμα δοθέντος ενός προβλήματος μη Επιβλεπόμενης Μάθησης. Το ζήτημα είναι ότι κάθε αλγόριθμος ακολουθεί μια διαφορετική μεθοδολογία, ορίζοντας με το δικό του τρόπο την «ομοιότητα» μεταξύ των δειγμάτων του συνόλου δεδομένων [79]. Έτσι, μπορούν να προκύψουν πολλά σωστά αποτελέσματα, αν και διαφορετικά μεταξύ τους, ανάλογα με τον αλγόριθμο που θα χρησιμοποιηθεί. Συνεπώς, η επιλογή θα πρέπει να γίνει ανάλογα με εκείνο που είναι επιθυμητό κάθε φορά, έχοντας φροντίσει πρώτα για τη «σωστή» επιλογή χαρακτηριστικών, χωρίς πλεονάζουσα πληροφορία, το μέτρο εγγύτητας που θα χρησιμοποιηθεί, καθώς και το κατάλληλο κριτήριο ομαδοποίησης, ανάλογα με τον τύπο / μορφή των συστάδων [73]. Όσον αφορά τους αλγόριθμους συσταδοποίησης, υπάρχει

μεγάλη ποικιλία, κάτι που μας επιτρέπει να διαλέγουμε τον καταλληλότερο ανάλογα με το εκάστοτε πρόβλημα και τους στόχους μας.

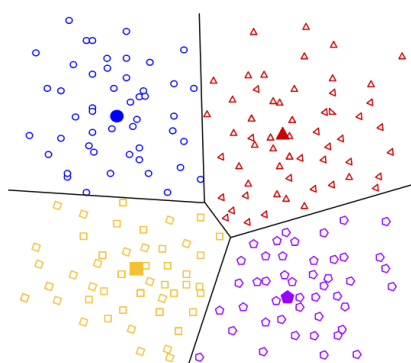
### 3.5.1 Αλγόριθμοι Συσταδοποίησης

Οι αλγόριθμοι συσταδοποίησης μπορούν να χωριστούν στις παρακάτω κύριες κατηγορίες:

- **Διαχωριστικοί αλγόριθμοι Συσταδοποίησης (Partitional Clustering)**

Οι αλγόριθμοι που ανήκουν σε αυτήν την κατηγορία διαχωρίζουν τα δεδομένα σε μη επικαλυπτόμενες συστάδες (non-overlapping clusters), με τέτοιο τρόπο ώστε κάθε στοιχείο να ανήκει ακριβώς σε μία ομάδα. Ο διαμερισμός των σημείων σε συστάδες γίνεται με στόχο να βελτιστοποιηθεί κάποια αντικειμενική συνάρτηση (objective function) που έχει τεθεί, όπως για παράδειγμα η μείωση του τετραγωνικού σφάλματος. Σε αυτόν τον τύπο συσταδοποίησης συγκαταλέγονται οι αλγόριθμοι με *βάση τα κέντρα* (centroid-based), διασημότερος εκ των οποίων είναι ο k-means, και εκείνοι με *βάση τα εσωτερικά κέντρα* (medoid-based).

Οι διαχωριστικοί αλγόριθμοι είναι σχετικά ευέλικτοι και απλοί στην υλοποίηση, κατάλληλοι για μεγάλα σύνολα δεδομένων και ακριβείς στην περίπτωση συμπαγών σφαιρικών συστάδων, οι οποίες είναι καλά διαχωρίσιμες. Ωστόσο, είναι ευαίσθητοι στη φάση αρχικοποίησης, στο θόρυβο και τα ακραία σημεία (outliers), ενώ επίσης απαιτείται η εκ των προτέρων γνώση του αριθμού συστάδων [80].



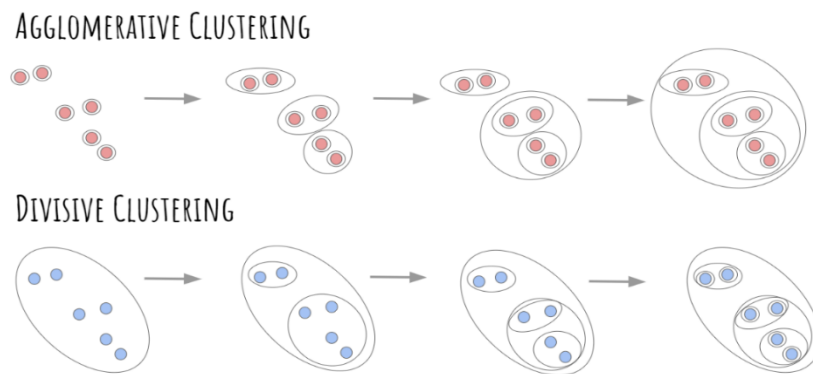
Σχήμα 3.5: Παράδειγμα συσταδοποίησης με χρήση διαχωριστικών αλγορίθμων. (Πηγή [81])

- **Ιεραρχικοί αλγόριθμοι Συσταδοποίησης (Hierarchical Clustering)**

Στόχος των αλγορίθμων που ανήκουν σε αυτήν την κατηγορία είναι να δημιουργήσουν εμφωλευμένες (nested) συστάδες, μέσω διαδοχικών συγχωνεύσεων (merges) ή διαχωρισμών (splits) τους. Ακολουθείται μια ιεραρχική

δομή, η οποία μπορεί να αναπαρασταθεί με τη βοήθεια δενδρογράμματος. Κάθε συστάδα περιλαμβάνει, τελικά, υποομάδες οργανωμένες σε ένα ιεραρχικό δέντρο [76]. Υπάρχουν δύο βασικοί τύποι Ιεραρχικής Συσταδοποίησης:

- Συσσωρευτικοί αλγόριθμοι (Agglomerative algorithms): ξεκινούν θεωρώντας ότι κάθε σημείο αποτελεί μία διαφορετική συστάδα και παράγουν μία ακολουθία ομαδοποιήσεων, όπου σε κάθε βήμα οι δύο κοντινότερες συστάδες συγχωνεύονται σε μία, μειώνοντας έτσι σταδιακά τον αριθμό των συστάδων (bottom-up approach). Οι αντιπροσωπευτικότεροι συσσωρευτικοί αλγόριθμοι είναι οι *απλού δεσμού* (single linkage) και οι *πλήρους δεσμού* (complete linkage) [73].
- Διαιρετικοί αλγόριθμοι (Divisive algorithms): ξεκινούν θεωρώντας ότι όλα τα σημεία ανήκουν σε μία μόνο συστάδα και παράγουν μία ακολουθία ομαδοποιήσεων, όπου σε κάθε βήμα κάθε συστάδα χωρίζεται σε δύο μικρότερες και έτσι αυξάνεται σταδιακά ο αριθμός των συστάδων (top-down approach).



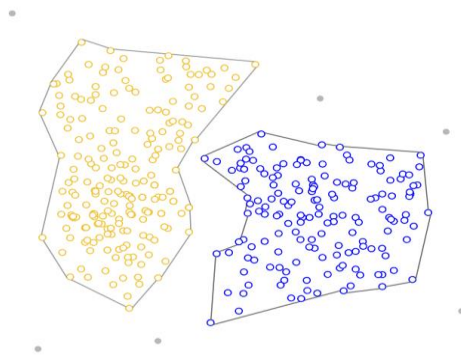
Σχήμα 3.6: Διαδικασία Ιεραρχικής Συσταδοποίησης με συσσωρευτικούς (πάνω) και διαιρετικούς (κάτω) αλγόριθμους. (Πηγή [82])

Οι Ιεραρχικοί αλγόριθμοι είναι χρήσιμοι όταν δε γνωρίζουμε τον ακριβή αριθμό συστάδων, κάτι που τους καθιστά αρκετά ευέλικτους, ως προς το επίπεδο λεπτομέρειας του συνόλου δεδομένων [76]. Όμως, παρουσιάζουν αδυναμία διόρθωσης πιθανών σφαλμάτων κατά τη διαδικασία συσταδοποίησης [80], οπότε αφού γίνει η εκάστοτε συγχώνευση, ή αντίστοιχα διαχωρισμός, δεν μπορεί να αλλάξει, παρά μόνο εάν ξεκινήσει η διαδικασία από την αρχή, κάτι που οδηγεί σε σπατάλη χρόνου και υπολογιστικών πόρων.

- **Αλγόριθμοι με βάση την πυκνότητα (Density-based Clustering)**

Αυτοί οι αλγόριθμοι ομαδοποίησης αναζητούν περιοχές στο χώρο δεδομένων, όπου υπάρχει μεγάλη πυκνότητα δειγμάτων και έτσι δημιουργούν τις συστάδες. Οι περισσότεροι από αυτούς απαιτούν λίγα περάσματα πάνω στο εκάστοτε σύνολο δεδομένων, ενώ διαφοροποιούνται μεταξύ τους ανάλογα με τον τρόπο με τον οποίο ποσοτικοποιούν τον όρο πυκνότητα [73].

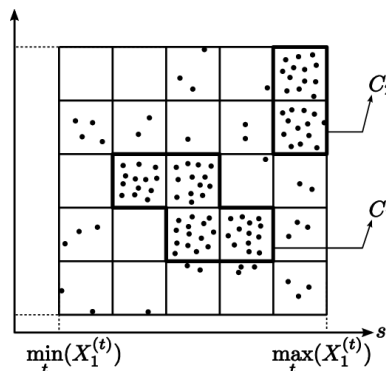
Είναι αποτελεσματικοί στην ανακάλυψη συστάδων αυθαίρετου σχήματος και διαφορετικών μεγεθών, ενώ επίσης είναι «ανθεκτικοί» στο θόρυβο και τα ακραία σημεία (outliers). Ωστόσο, βρίσκουν δυσκολία στα σύνολα δεδομένων πολλών διαστάσεων και είναι ευαίσθητοι όσον αφορά τις παραμέτρους εισόδου (input parameters) [80]. Ευρέως χρησιμοποιούμενοι αλγόριθμοι της κατηγορίας αυτής είναι οι DBSCAN και OPTICS.



Σχήμα 3.7: Παράδειγμα συσταδοποίησης με βάση την πυκνότητα. (Πηγή [81])

- **Αλγόριθμοι με βάση τα πλέγματα (Grid-based clustering)**

Σε αυτήν την κατηγορία ανήκουν οι αλγόριθμοι που χρησιμοποιούν δομή πολυεπίδεδου πλέγματος (multi resolution grid) για να σχηματίσουν τις συστάδες. Πιο συγκεκριμένα, χωρίζουν το χώρο δεδομένων σε πεπερασμένο αριθμό κελιών, δημιουργώντας την προαναφερθείσα δομή πλέγματος, πάνω στην οποία και εκτελείται η διαδικασία της συσταδοποίησης [80].



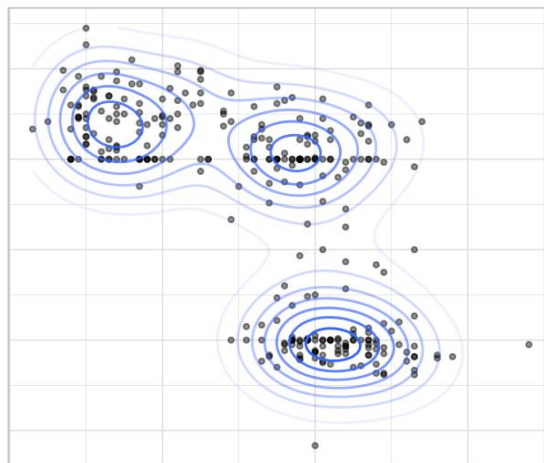
Σχήμα 3.8: Παράδειγμα συσταδοποίησης με βάση τα πλέγματα. (Πηγή [83])

Το βασικό πλεονέκτημα αυτών των μοντέλων είναι η ταχύτητα στην επεξεργασία των δεδομένων, ακόμη και εάν πρόκειται για μεγάλα σύνολα. Παρόλα αυτά η επιλογή του κατάλληλου μεγέθους πλέγματος είναι πολύ σημαντική, διαφορετικά ενδέχεται να προκύψουν συστάδες με πάρα πολλά ή αντίστοιχα λίγα δείγματα, επηρεάζοντας αρνητικά την ποιότητα της συσταδοποίησης.

- **Αλγόριθμοι με βάση μοντέλα (Model-based clustering)**

Οι αλγόριθμοι αυτοί θεωρούν ότι τα δεδομένα προέρχονται από μια μίξη πεπερασμένου αριθμού μαθηματικών μοντέλων. Κάθε μοντέλο αποτελεί στην πραγματικότητα μια κατανομή πιθανότητας (probability distribution), όπως είναι για παράδειγμα η Γκαουσιανή κατανομή [84]. Υποθέτουν ένα μοντέλο για κάθε συστάδα και στόχος τους είναι να ανακαλύψουν το καλύτερο δυνατό «ταίριασμα» μεταξύ δεδομένων – μοντέλων, ώστε να αντιστοιχίσουν κάθε σημείο του χώρου δεδομένων στην κατάλληλη συστάδα, δηλαδή σε εκείνη στην οποία αναμένεται να έχει τη μεγαλύτερη πιθανότητα να ανήκει. Να σημειωθεί, ωστόσο, ότι πέραν της στατιστικής αυτής προσέγγισης σε αυτήν την κατηγορία συσταδοποίησης ανήκει και η μέθοδος εκπαίδευσης μέσω των νευρωνικών δικτύων, με χαρακτηριστικά παραδείγματα τους αλγόριθμους SOM και ART [85].

Σε αυτό το είδος συσταδοποίησης παρατηρούνται ποικίλα σχήματα συστάδων, συχνά με περίπλοκες δομές. Επιπλέον, εμφανίζει το πλεονέκτημα του ότι αυτομάτως εντοπίζει τον ιδανικό αριθμό συστάδων, χωρίς να χρειάζεται να προκαθοριστεί. Ωστόσο, οι αλγόριθμοι αυτής της κατηγορίας έχουν υψηλή χρονική πολυπλοκότητα (time complexity) και τείνουν να είναι ευαίσθητοι στις παραμέτρους των επιλεγμένων μοντέλων [85].



Σχήμα 3.9: Παράδειγμα συσταδοποίησης με βάση μοντέλα. (Πηγή [86])

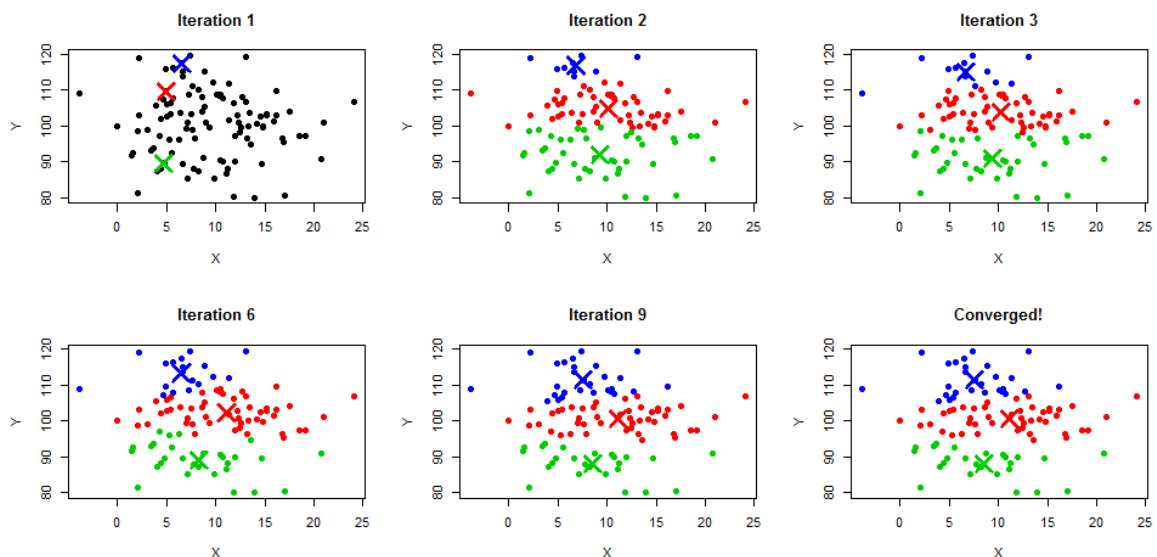
Παρακάτω αναλύεται ο ευρέως χρησιμοποιούμενος αλγόριθμος k-Μέσων (k-Means), ο οποίος χρησιμοποιείται και στο πειραματικό κομμάτι της εργασίας.

### 3.5.1.1 Αλγόριθμος k-Means

Πρόκειται για έναν από τους πιο δημοφιλείς αλγορίθμους συσταδοποίησης και ανήκει στην κατηγορία του *partitional clustering*. Ομαδοποιεί τα δεδομένα σε  $k$  συστάδες και αυτό το πετυχαίνει ελαχιστοποιώντας μια συνάρτηση που βασίζεται στο άθροισμα των τετραγώνων των σφαλμάτων (*within-cluster sum-of-squares criterion*) ή αλλιώς γνωστή ως κριτήριο αδράνειας (*inertia*).

Ο αλγόριθμος k-means ξεκινά με τυχαία αρχικοποίηση  $k$  σημείων στο χώρο δεδομένων, τα οποία θεωρεί ως κέντρα (*centroids*) των  $k$  συστάδων που πρόκειται να δημιουργηθούν. Στη συνέχεια, εκτελεί τα παρακάτω δύο βήματα σε επανάληψη [65]:

1. Ανάθεση των παρατηρήσεων στη συστάδα με το κοντινότερο κέντρο: *βασίζόμενος στην τιμή του κριτηρίου inertia εξετάζει την απόσταση δείγματος -κέντρου και το κατηγοριοποιεί στην ομάδα με το πλησιέστερο προς το δείγμα κέντρο.*
2. Εκ νέου προσδιορισμός των κέντρων κάθε συστάδας με βάση την τρέχουσα κατανομή των παρατηρήσεων: *υπολογίζεται η μέση τιμή των δειγμάτων κάθε συστάδας και με αυτόν τον τρόπο επαναπροσδιορίζεται το κέντρο της, ώστε να αντικατοπτρίζει το νέο «κέντρο βάρους» των παρατηρήσεων βάσει της πρόσφατα διαμορφωμένης ομάδας.*



Σχήμα 3.10: Εκτέλεση του αλγορίθμου k-means με  $k=3$ . (Πηγή [87])

Ουσιαστικά, ο αλγόριθμος αυτός ακολουθεί μια άπληστη τεχνική, η οποία τερματίζει όταν τα κέντρα των συστάδων μετατοπίζονται πια ελάχιστα, δηλαδή η απόσταση μεταξύ παλιών και νέων centroids είναι μικρότερη από κάποια ορισμένη τιμή κατωφλίου ή όταν έχει καθοριστεί συγκεκριμένος αριθμός επαναλήψεων για τον αλγόριθμο. Σε κάθε περίπτωση, ο k-means συγκλίνει, όμως τα αποτελέσματα που προκύπτουν εξαρτώνται σε μεγάλο βαθμό από την αρχικοποίηση των κέντρων. Για να αποφευχθεί μια «κακή» αρχικοποίηση που θα επηρεάσει αρνητικά τη διαδικασία συσταδοποίησης μπορεί να χρησιμοποιηθεί η παράμετρος *k-means++*, εξασφαλίζοντας τουλάχιστον ότι τα αρχικά κέντρα θα είναι σχετικά απομακρυσμένα μεταξύ τους, κάτι το οποίο τείνει να οδηγεί σε καλύτερα αποτελέσματα [88].

Συμπερασματικά, ο k-means αποτελεί έναν απλό και εύκολα υλοποιήσιμο αλγόριθμο με πολυπλοκότητα  $O(kNt)$ , όπου  $k$  το πλήθος συστάδων,  $N$  το πλήθος των δεδομένων και  $t$  ο αριθμός των επαναλήψεων (iterations). Το γεγονός αυτό τον καθιστά αρκετά αποδοτικό ακόμη και για μεγάλα σύνολα δεδομένων. Επιπλέον, δίνει συμπαγείς και καλά διαχωρίσιμες συστάδες, βρίσκοντας εφαρμογή σε πληθώρα προβλημάτων. Ωστόσο, απαιτεί τον καθορισμό ακριβή αριθμού  $k$  συστάδων, χωρίς να υπάρχει κάποια αυτοματοποιημένη μέθοδος επιλογής του που να λειτουργεί σωστά σε κάθε πρόβλημα. Ορισμένοι συχνά χρησιμοποιούμενοι αλγόριθμοι για τον προσδιορισμό του  $k$  είναι η μέθοδος του αγκώνα (*elbow method*) και η μέθοδος «silhouette» (*silhouette score method*). Ένα ακόμη σημαντικό μειονέκτημά του είναι η ευαισθησία του στο θόρυβο και στα ακραία σημεία (outliers), λόγω του ότι πρέπει να κατηγοριοποιηθούν όλες οι παρατηρήσεις ανεξαρτήτως της απόστασής τους από τα centroids, όπως επίσης και το ότι δε δουλεύει για μη αριθμητικά δεδομένα και δεν μπορεί να εγγυηθεί εύρεση βέλτιστης λύσης [89].

### 3.5.2 Εγκυρότητα συσταδοποίησης

Ως εγκυρότητα συσταδοποίησης (cluster validity) ορίζεται η διαδικασία ποσοτικής αξιολόγησης των αποτελεσμάτων ενός αλγορίθμου συσταδοποίησης [73]. Αποτελεί ένα πολύ σημαντικό κομμάτι του τομέα της ανάλυσης σε συστάδες (cluster analysis), ένα χρήσιμο εργαλείο για την ανατροφοδότηση του ειδικού ως προς την επίδοση των αλγορίθμων. Ένα από τα βασικά προβλήματα της διαδικασίας, ωστόσο, αποτελεί το γεγονός ότι δεν υπάρχει σαφής έννοια για το σφάλμα πρόβλεψης (prediction error), όπως στην επιβλεπόμενη μάθηση. Αυτό έχει σαν συνέπεια, οι αλγόριθμοι αυτοί να αξιολογούνται βάσει ποικίλων, συχνά αυθαίρετων, μετρικών που ενδέχεται να μην είναι απολύτως αντιπροσωπευτικές [90]. Έτσι, η αξιολόγηση προβλημάτων συσταδοποίησης



αποδεικνύεται, ορισμένες φορές, πολύπλοκη διαδικασία, ιδίως όταν εμπλέκονται πολλές διαστάσεις. Υπάρχουν τρεις διαφορετικές προσεγγίσεις για την επαλήθευση της εγκυρότητας των προβλημάτων αυτών, τα εξωτερικά, τα εσωτερικά και τα σχετικά κριτήρια.

Σύμφωνα με τα *εξωτερικά κριτήρια εγκυρότητας* (external clustering validity methods), χρησιμοποιείται προϋπάρχουσα γνώση, ανεξάρτητη της εν λόγω συσταδοποίησης, όπως είναι για παράδειγμα οι πραγματικές ετικέτες των δειγμάτων (ground truth labels), ώστε συγκρίνοντάς τες με τις ετικέτες που προβλέπει ο εκάστοτε αλγόριθμος (predicted labels) να βοηθήσουν στην αξιολόγηση. Τέτοια κριτήρια είναι οι μετρικές που ανήκουν στην οικογένεια των Mutual information scores (MI, NMI, AMI) και η Rand index. Η μεθοδολογία που ακολουθούν περιλαμβάνει τη σύγκριση των πραγματικών ετικετών με αυτών που προκύπτουν μετά τη συσταδοποίηση. Ωστόσο, έχει αναφερθεί ότι τα εξωτερικά κριτήρια δεν αποτελούν τη βέλτιστη επιλογή για πραγματικές (real-world) εφαρμογές, αλλά προτιμώνται για συνθετικά δεδομένα [91].

Σύμφωνα με τα *εσωτερικά κριτήρια εγκυρότητας* (internal clustering validity methods), για την αξιολόγηση της συσταδοποίησης χρησιμοποιούνται μόνο τα ίδια τα δεδομένα, χωρίς να εμπλέκονται «εξωτερική» γνώση. Τέτοια κριτήρια είναι το Silhouette score, ο δείκτης Davies–Bouldin, καθώς και ο Calinski–Harabasz. Υπολογίζουν τη συνεκτικότητα εντός συστάδας (within-cluster cohesion) και το διαχωρισμό μεταξύ των συστάδων (between-cluster separation) [90] και με αυτόν τον τρόπο αξιολογούν το clustering. Αποτελούν συχνά εφαρμοζόμενες μεθόδους, έχοντας πολλές πρακτικές εφαρμογές.

Τέλος, σύμφωνα με τα *σχετικά κριτήρια εγκυρότητας* (relative clustering validity methods), η αξιολόγηση της συσταδοποίησης προκύπτει από τη σύγκρισή της με άλλες ομαδοποιήσεις που είτε έχουν χρησιμοποιήσει διαφορετικό αλγόριθμο, είτε τον ίδιο με διαφορετικές παραμέτρους κάθε φορά [73]. Τα κριτήρια αυτά συχνά χρησιμοποιούνται για τον καθορισμό του βέλτιστου αριθμού  $k$  των συστάδων [90].

### 3.6 Νευρωνικά Δίκτυα

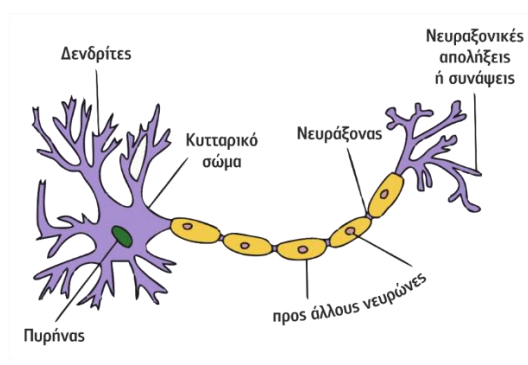
Στο παρόν κεφάλαιο, θα ασχοληθούμε με τον κλάδο των Νευρωνικών Δικτύων και την καθοριστική συμβολή τους στην ανάπτυξη προηγμένων μοντέλων μηχανικής μάθησης. Ιστορικά, ο όρος νευρωνικά δίκτυα αναφέρεται σε δίκτυα νευρώνων που σχηματίζονται στον εγκέφαλο των θηλαστικών και είναι γνωστά ως Βιολογικά Νευρωνικά Δίκτυα. Οι νευρώνες αυτοί αποτελούν τις θεμελιώδεις μονάδες υπολογισμού του εγκεφάλου, οι οποίες

συνδέονται μεταξύ τους δημιουργώντας δίκτυα με σκοπό την επεξεργασία δεδομένων [92]. Ωστόσο, πλέον, με τον όρο «Νευρωνικά Δίκτυα», συνήθως δεν αναφερόμαστε στα Βιολογικά, αλλά στα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ), τα οποία αποτελούν υποσύνολο του κλάδου της Μηχανικής Μάθησης και έχουν γνωρίσει μεγάλη δημοσιότητα τα τελευταία χρόνια, βελτιώνοντας τις επιδόσεις σε πολλούς τομείς της επιστήμης και της τεχνολογίας. Πρόκειται για ένα σύνολο ψηφιακών νευρώνων, συνδεδεμένων μεταξύ τους με συγκεκριμένη αρχιτεκτονική δομή, που στόχο έχει τη μοντελοποίηση του τρόπου με τον οποίο ο εγκέφαλος πραγματοποιεί κάποια εργασία ή κάποια λειτουργία [75].

Παρακάτω αναλύεται περιληπτικά η δομή και ο τρόπος λειτουργίας των Βιολογικών Νευρωνικών Δικτύων, τα οποία υπήρξαν πηγή έμπνευσης για τη δημιουργία των ΤΝΔ, με τα οποία θα ασχοληθούμε ευρέως στην παρούσα εργασία, ξεκινώντας με την ιστορία τους, παρουσιάζοντας ορισμένα πλεονεκτήματά τους και εστιάζοντας σε δημοφιλείς αρχιτεκτονικές τους.

### Βιολογικά Νευρωνικά Δίκτυα

Τα συστατικά στοιχεία ενός Βιολογικού Νευρωνικού Δικτύου (Biological Neural Network) είναι οι νευρώνες μαζί με τις διασυνδέσεις τους. Οι νευρώνες (neurons) είναι κύτταρα του νευρικού ιστού και αποτελούν τη βασική μονάδα δόμησης του εγκέφαλου. Αποτελούνται από 3 κύρια τμήματα:



Σχήμα 3.11: Δομή ενός βιολογικού νευρώνα. (Πηγή [61])

- α) τους **δενδρίτες** (dendrites), έναν αριθμό από ίνες, οι οποίες λειτουργούν ως κανάλια εισόδου της πληροφορίας,
- β) το κυρίως **κυτταρικό σώμα** (cell body / soma) και
- γ) τον **άξονα** του κυττάρου ή αλλιώς νευροάξονα (axon), μια επιμήκη ίνα που είναι υπεύθυνη για τη σύνδεση των νευρώνων μεταξύ τους [61].

Ο ανθρώπινος εγκέφαλος αποτελείται από έναν πολύ μεγάλο αριθμό νευρώνων (της τάξης του  $10^{10}$ ), ενώ καθένας από αυτούς συνδέεται με πολλούς άλλους. Οι συνδέσεις αυτές ονομάζονται συνάψεις (synapses) και συνήθως ένας νευρώνας κάνει περίπου  $10^4$  συνάψεις, καθ' όλη τη διάρκεια της ζωής του οργανισμού. Η δημιουργία των νέων συνάψεων, ωστόσο, δε γίνεται σε κάποια τυχαία χρονική στιγμή της ζωής του, αλλά όταν ο εγκέφαλος αποκτά εμπειρίες από το περιβάλλον του, μαθαίνει, αναγνωρίζει και κατανοεί [93].

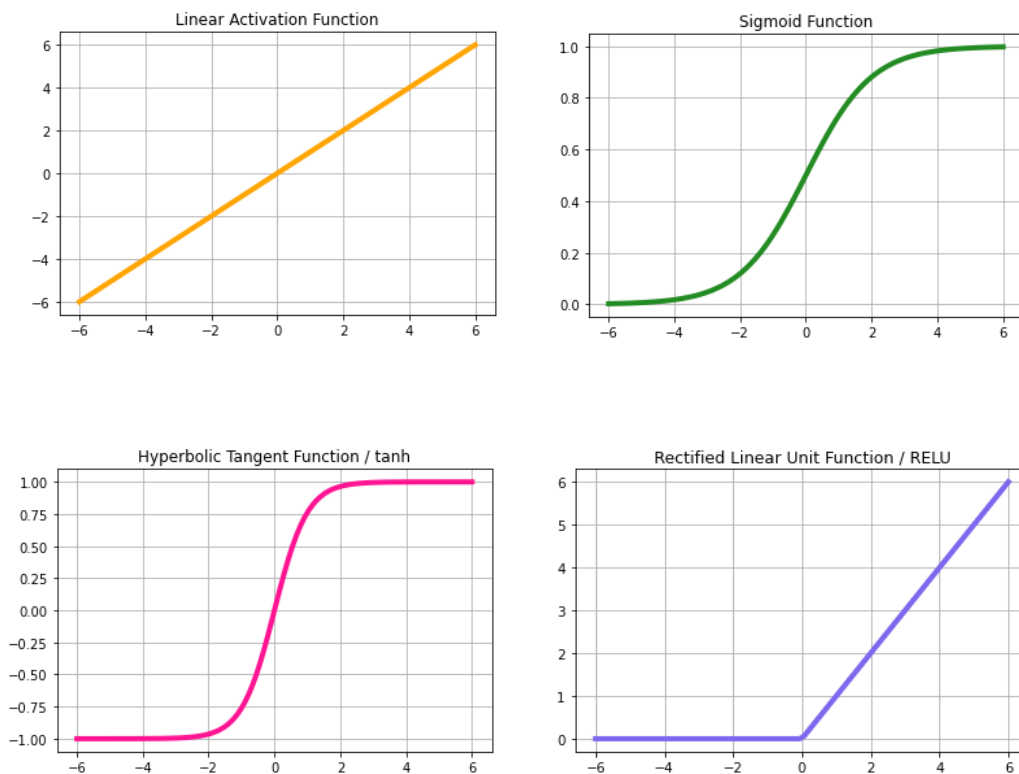
Συνεπώς, γίνεται αντιληπτό ότι ο εγκέφαλος έχει μια εξαιρετικά πολύπλοκη δομή που, όμως, έχει τη δυνατότητα να οργανώνει σε ομάδες, τα νευρωνικά δίκτυα, τα οποία «δουλεύουν» παράλληλα, με τρόπο ώστε να εκτελούν συγκεκριμένες λειτουργίες. Για να συμβεί αυτό, οι νευρώνες επικοινωνούν μεταξύ τους, λαμβάνοντας και μεταδίδοντας ηλεκτρικά σήματα (μερικών mV), αφού πρώτα τα επεξεργαστούν. Πιο συγκεκριμένα, υπάρχουν δύο καταστάσεις στις οποίες δύναται να βρεθεί ο κάθε νευρώνας: η ενεργός και η μη-ενεργός. Όταν βρίσκεται σε ενεργό κατάσταση, τότε παράγει ηλεκτρικό σήμα, ενώ όταν είναι σε μη-ενεργό θεωρείται αδρανής. Η κατάσταση του προσδιορίζεται από τα σήματα που λαμβάνει σε κάθε χρονική στιγμή. Τα ηλεκτρικά δυναμικά αυτών των σημάτων αθροίζονται και εάν το αποτέλεσμα προκύπτει μεγαλύτερο ή ίσο μιας συγκεκριμένης τιμής (τιμή κατωφλίου), τότε λέμε ότι βρίσκεται σε διεγερμένη κατάσταση και στέλνει μέσω του άξονά του το σήμα, ενώ εάν προκύπτει μικρότερο, τότε παραμένει μη-ενεργός και το σήμα χάνεται. Σε αυτήν την αρχή λειτουργίας των Βιολογικών Νευρωνικών Δικτύων στηρίζονται τα ΤΝΔ, όπως εξηγείται παρακάτω.

### 3.6.1 Τεχνητά Νευρωνικά Δίκτυα

Πρόκειται για ένα σύνολο μαθηματικών μοντέλων που βασίζονται στη λειτουργία του ανθρώπινου εγκεφάλου. Στην ουσία, ως Τεχνητό Νευρωνικό Δίκτυο - ΤΝΔ (Artificial Neural Networks - ANN) θα μπορούσε να οριστεί ένα ευρύ δίκτυο με κατανομημένη αρχιτεκτονική, το οποίο επεξεργάζεται παράλληλα τις πληροφορίες που δέχεται και αποτελείται από απλές μονάδες επεξεργασίας, τους τεχνητούς νευρώνες. Έχει τη δυνατότητα να αποθηκεύει εμπειρική γνώση, την οποία αποκτά από το περιβάλλον του μέσω μιας διαδικασίας μάθησης [75] και να γενικεύει, δηλαδή να την χρησιμοποιεί σε περιπτώσεις που δεν έχει συναντήσει ξανά [94].

Όσον αφορά τη δομή τους, αποτελούνται από απλές, αλλά πολύπλοκα διασυνδεδεμένες επεξεργαστικές μονάδες, τους νευρώνες (neurons). Κάθε τεχνητός νευρώνας, σε αναλογία με τους βιολογικούς, διαθέτει έναν αριθμό συνάψεων, με τη

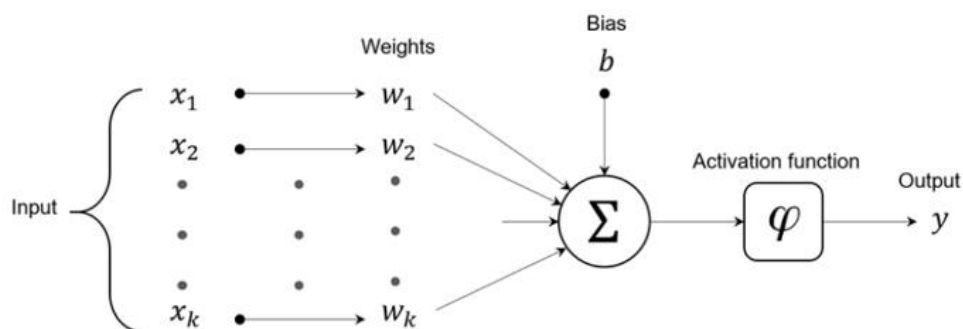
βοήθεια των οποίων συνδέεται και επικοινωνεί με τους υπόλοιπους νευρώνες του δικτύου, στέλνοντας και λαμβάνοντας σήματα. Οι συνάψεις αυτές χαρακτηρίζονται από βάρη  $w$  (weights), τα οποία καθορίζονται από την ισχύ της σύνδεσης μεταξύ των εκάστοτε νευρώνων. Καλούνται συναπτικά βάρη και επηρεάζουν άμεσα τη διαδικασία της μάθησης. Βάσει αυτών των βαρών «σταθμίζονται» τα σήματα στην είσοδο της κάθε συνάψης, δηλαδή πολλαπλασιάζονται με τους εκάστοτε συντελεστές βαρύτητας. Τα αποτελέσματα καταλήγουν σε έναν αθροιστή (adder), ο οποίος υπολογίζει το σταθμισμένο άθροισμα (weighted sum) αυτών των σημάτων εισόδου. Τελικά, η τιμή που προκύπτει χρησιμοποιείται σαν είσοδος σε μια συνάρτηση ενεργοποίησης  $\varphi$  (activation function), η οποία περιορίζει το πλάτος του σήματος εξόδου σε κάποια πεπερασμένη τιμή [75], αποφασίζοντας, στην ουσία, εάν θα ενεργοποιηθεί ή όχι ο εν λόγω νευρώνας. Ο ρόλος της είναι να προσθέσει μη-γραμμικότητα στην έξοδό του. Υπάρχουν τρεις τύποι συναρτήσεων ενεργοποίησης: α) οι συναρτήσεις απότομου περιορισμού (hard-limit functions), δηλαδή η βηματική και η συνάρτηση προσήμου, β) η γραμμική συνάρτηση (linear function) και γ) οι μη-γραμμικές συναρτήσεις (non-linear functions). Στα σύγχρονα ΤΝΔ ευρέως χρησιμοποιείται η γραμμική συνάρτηση ενεργοποίησης (linear activation function), ενώ από τις μη-γραμμικές, η σιγμοειδής (sigmoid), η υπερβολική εφαπτομένη (tanh), καθώς και η ανορθωμένη γραμμική μονάδα (ReLU).



Σχήμα 3.12: Συναρτήσεις ενεργοποίησης.

Ένα ακόμη συστατικό στοιχείο του τεχνητού νευρώνα αποτελεί η πόλωση  $b$  (bias), η οποία μπορεί να θεωρηθεί ως μια εξωτερικά εφαρμοζόμενη σταθερά που έχει σαν αποτέλεσμα την αύξηση ή μείωση της δικτυακής διέγερσης της συνάρτησης ενεργοποίησης [75]. Με απλά λόγια, επιτρέπει τη «μετατόπιση» (shift) της συνάρτησης ενεργοποίησης προς τα δεξιά ή αριστερά, κάτι που έχει σαν συνέπεια τη διαφοροποίηση του τελικού αποτελέσματος, της εξόδου. Εφαρμόζεται σαν είσοδος στον αθροιστή που προαναφέρθηκε, οπότε προστίθεται στα γινόμενα των συναπτικών βαρών με τις εκάστοτε εισόδους και θεωρείται εξαιρετικά χρήσιμη, εφόσον βοηθά το μοντέλο, ώστε να γίνει αποτελεσματικότερα η εκπαίδευση.

Να σημειωθεί ότι κάθε νευρώνας λαμβάνει αρκετά σήματα εισόδου μέσω των συνάψεών του, που προέρχονται είτε από ακατέργαστα δεδομένα (raw data) απευθείας, είτε από άλλους νευρώνες. Αφού επεξεργαστεί τις πληροφορίες που δέχτηκε με τον τρόπο που περιγράφηκε παραπάνω, παράγει ένα μόνο σήμα εξόδου, το οποίο, με τη σειρά του, διαδίδεται σε γειτονικούς νευρώνες ή αποτελεί την τελική έξοδο στο πρόβλημα.



Σχήμα 3.13: Δομή ενός τεχνητού νευρώνα. (Πηγή [95])

Η εξίσωση που δίνει την έξοδο ενός νευρώνα και περιγράφει με μαθηματικούς όρους τη δομή ενός τεχνητού νευρώνα θα μπορούσε να γραφεί ως εξής:

$$y = \varphi\left(\sum_{i=1}^k w_i x_i + b\right)$$

όπου  $x_i$  είναι τα σήματα εισόδου,  $w_i$  τα αντίστοιχα συναπτικά βάρη για τον συγκεκριμένο νευρώνα,  $b$  η τιμή της πόλωσης,  $\varphi(\bullet)$  η συνάρτηση ενεργοποίησης και  $y$  η τιμή της εξόδου του νευρώνα.

Ιστορικά, πρώτοι, οι οποίοι αναφέρθηκαν στον όρο αυτό ήταν οι Warren McCulloch και Walter Pitts, το 1943 [96], περιγράφοντας τον τρόπο με τον οποίο θα μπορούσαν να λειτουργούν οι νευρώνες και μοντελοποιώντας ένα απλό ΤΝΔ με τη βοήθεια ηλεκτρικών κυκλωμάτων. Και άλλες προσπάθειες έγιναν μέχρι το 1957, οι οποίες όμως δεν είχαν επιτυχία, μέχρι που ένας νευροβιολόγος, ο Frank Rosenblatt προτείνει ένα ολοκληρωμένο μοντέλο νευρώνα, το Perceptron [97], που θεωρείται ως το παλαιότερο ΤΝΔ που χρησιμοποιείται μέχρι και σήμερα. Στα χρόνια, ωστόσο, που ακολούθησαν οι έρευνες στον κλάδο των ΤΝΔ πάγωσαν, λόγω του ότι, καταρχάς, η τεχνολογία δε διευκόλυνε την κατάσταση και παράλληλα η μοντελοποίηση αυτών των δικτύων ήταν αρκετά κοστοβόρα, ενώ επίσης, είχαν δημοσιευθεί έρευνες που αποθάρρυναν τους επιστήμονες σε σχέση με την ενασχόληση με τα Νευρωνικά Δίκτυα αποδεικνύοντας μαθηματικά τους θεμελιώδεις περιορισμούς των υπολογιστικών δυνατοτήτων του Perceptron και την αδυναμία του στην επίλυση μη γραμμικών διαχωριζόμενων προβλημάτων. Διέξοδος στο τέλμα στο οποίο είχαν οδηγηθεί οι ειδήμονες της Τεχνητής Νοημοσύνης αποτέλεσαν η δημοσίευση του Hopfield το 1982 [98], ο οποίος διατύπωσε τη φυσική αρχή της αποθήκευσης πληροφορίας σε ένα δυναμικά ευσταθές δίκτυο και παρουσίασε μια κατηγορία νευρωνικών δικτύων με ανάδραση (δίκτυα Hopfield), καθώς και το βιβλίο των Rumelhart και McClelland το 1986 [95], όπου παρουσίασαν τον αλγόριθμο μάθησης με οπισθοδιάδοση σφάλματος (back-propagation learning algorithm). Αυτά ίσως αποτέλεσαν τα δύο σημαντικότερα έργα στα οποία οφείλεται η αναγέννηση των ΤΝΔ τη δεκαετία του 1980 [94].

Έκτοτε και μέχρι σήμερα, και με την έλευση της Βαθιάς Μάθησης κατά τη δεκαετία του 2000, ο κλάδος των ΤΝΔ έχει γνωρίσει αλματώδη ανάπτυξη. Σε αυτό σίγουρα συνέβαλε η εκρηκτική εξέλιξη του Διαδικτύου με το δυσθεώρητο όγκο πληροφοριών που αυτό παρέχει, καθώς κατέστησε σημαντικά ευκολότερη την πρόσβαση σε πληθώρα δεδομένων και επέτρεψε την προσέγγιση προβλημάτων που πρωτότερα δεν ήταν εφικτή, δίνοντας σε πολλές περιπτώσεις state-of-the-art αποτελέσματα. Επιπλέον, καθοριστικό ρόλο έπαιξε η ραγδαία πρόοδος στον κλάδο της παράλληλης επεξεργασίας των δεδομένων, τα τελευταία 25 χρόνια, που έφερε σημαντική βελτίωση των Γραφικών Μονάδων Επεξεργασίας (GPU - Graphics Processing Unit). Όπως έχει αναλυθεί, τα ΤΝΔ λειτουργούν εκ φύσεως παράλληλα, οπότε με τη χρήση των «βελτιωμένων» GPUs, που αξιοποιούν τον παράλληλο προγραμματισμό, διαχειρίζονται μεγάλους όγκους δεδομένων και χαρακτηρίζονται από ικανότητα ανάπτυξης μεγάλης υπολογιστικής ισχύος, ενώ, ταυτοχρόνως, η ταχύτητα της εκπαίδευσης και κατ' επέκταση της μάθησης επιταχύνθηκε σε μεγάλο βαθμό.

Όλα αυτά οδήγησαν στην εύρεση σύγχρονων τεχνικών και πιο αποδοτικών αλγορίθμων στον τομέα των Νευρωνικών Δικτύων, τα οποία πλέον χρησιμοποιούνται σε μια πληθώρα εφαρμογών και δύνανται να λύσουν προβλήματα όπου οι παραδοσιακοί μέθοδοι αντιμετώπισης τους παρουσίαζαν δυσκολίες. Σημαντική είναι η συνεισφορά τους στον τομέα της Μηχανικής, της Ρομποτικής, της Όρασης υπολογιστών, των Τηλεπικοινωνιών, της Αεροδιαστημικής, της Αυτοκινητοβιομηχανίας και ακόμη περισσότερο στον τομέα της Ιατρικής, των Επιχειρήσεων, των Μεταφορών, της Οικονομίας, αλλά και της Εκπαίδευσης. Χαρακτηριστικά παραδείγματα εφαρμογών των ΤΝΔ αποτελούν η αναγνώριση χειρόγραφων χαρακτήρων, η αυτόματη μετάφραση, η αναγνώριση φωνής και προσώπου, η διάγνωση ασθενειών, η πρόβλεψη του καιρού, η εκπαίδευση ρομπότ, καθώς και πολλά ακόμη.

### Πλεονεκτήματα ΤΝΔ

Το μεγάλο ενδιαφέρον της επιστημονικής κοινότητας προς τον κλάδο των Τεχνητών Νευρωνικών Δικτύων δεν ήταν τυχαίο. Τα ΤΝΔ παρουσιάζουν ορισμένα πλεονεκτήματα που δύσκολα αμελούνται:

- Παράλληλη επεξεργασία δεδομένων → Οι νευρώνες των ΤΝΔ λειτουργούν εκ φύσεως παράλληλα, οπότε αυξάνεται η ταχύτητα του συστήματος.
- Γενίκευση → Αποκτούν εμπειρία από το περιβάλλον τους, την οποία μπορούν να χρησιμοποιήσουν σε δεδομένα που δεν έχουν συναντήσει ξανά, έτσι ώστε να κάνουν προβλέψεις.
- Μη γραμμικότητα → Αποτελεί μια σημαντική ιδιότητα των ΤΝΔ, καθώς τους δίνει τη δυνατότητα να μαθαίνουν και να μοντελοποιούν περίπλοκες και μη γραμμικές συσχετίσεις μεταξύ εισόδου – εξόδου που βρίσκει εφαρμογή σε real-world προβλήματα.
- Προσαρμοστικότητα → Σε αναλογία με τις μεταβολές που γίνονται στο περιβάλλον τους οι νευρώνες έχουν τη δυνατότητα να προσαρμόζουν τα συναπτικά βάρη τους.
- Ανοχή στα σφάλματα → Σε περίπτωση βλάβης στα ΤΝΔ, η απόδοσή τους μειώνεται σταδιακά και δεν εμποδίζει την παραγωγή αποτελέσματος.
- Ομοιομορφία ανάλυσης και σχεδίασης → Χρησιμοποιείται η ίδια σημειογραφία σε όλα τα πεδία εφαρμογής των νευρωνικών δικτύων [75].
- Αναλογία με νευροφυσιολογία του εγκεφάλου → Δίνει τη δυνατότητα τόσο σε νευροβιολόγους όσο και σε μηχανικούς να βρίσκουν συσχετίσεις μεταξύ

βιολογικών και τεχνητών νευρωνικών δικτύων, ανακαλύπτοντας νέες πτυχές σε δυσεπίλυτα προβλήματα.

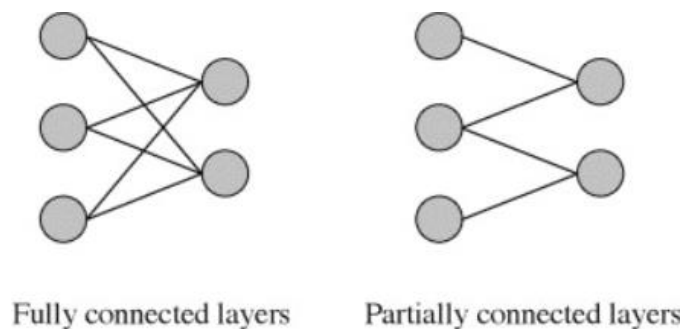
### Αρχιτεκτονικές ΤΝΔ

Ως αρχιτεκτονική ενός Τεχνητού Νευρωνικού Δικτύου ορίζεται ο τρόπος με τον οποίο είναι δομημένες οι απλές μονάδες επεξεργασίας του, δηλαδή οι νευρώνες και συνδέεται άμεσα με τη μέθοδο μάθησης, δηλαδή τον αλγόριθμο που θα χρησιμοποιηθεί για την εκπαίδευσή του [75].

Πριν εξετάσουμε, ωστόσο, τις κύριες αρχιτεκτονικές, βάσει των οποίων σχεδιάζονται τα δίκτυα αυτά, είναι σημαντικό να εξηγήσουμε ότι οι τεχνητοί νευρώνες οργανώνονται σε στρώματα (layers). Οπότε ένα τυπικό ΤΝΔ αποτελείται από μία ιεραρχία επιπέδων [94], το επίπεδο εισόδου, μέσω του οποίου τροφοδοτούμε με πληροφορίες το δίκτυο, τα κρυφά επίπεδα, όπου εκτελούνται όλοι οι σημαντικοί υπολογισμοί που είναι υπεύθυνοι για την εξαγωγή χαρακτηριστικών και κρυφών μοτίβων και το επίπεδο εξόδου, το οποίο αποδίδει τα τελικά αποτελέσματα.

Έχοντας αναλύσει τον τρόπο με τον οποίο οργανώνονται οι τεχνητοί νευρώνες «κάθετα», στα 3 διαφορετικά επίπεδα, μπορούμε να αναφερθούμε σε μια πρώτη διάκριση των ΤΝΔ ανάλογα με τον τρόπο σύνδεσης των νευρώνων τους. Χωρίζονται στις εξής δύο κατηγορίες:

- **Πλήρως συνδεδεμένα** (Fully connected), στα οποία κάθε νευρώνας συνδέεται με όλους τους νευρώνες - κόμβους του προηγούμενου επιπέδου.
- **Μερικώς συνδεδεμένα** (Partially connected), στα οποία υπάρχουν νευρώνες που δεν συνδέονται με όλους τους νευρώνες - κόμβους του προηγούμενου επιπέδου.

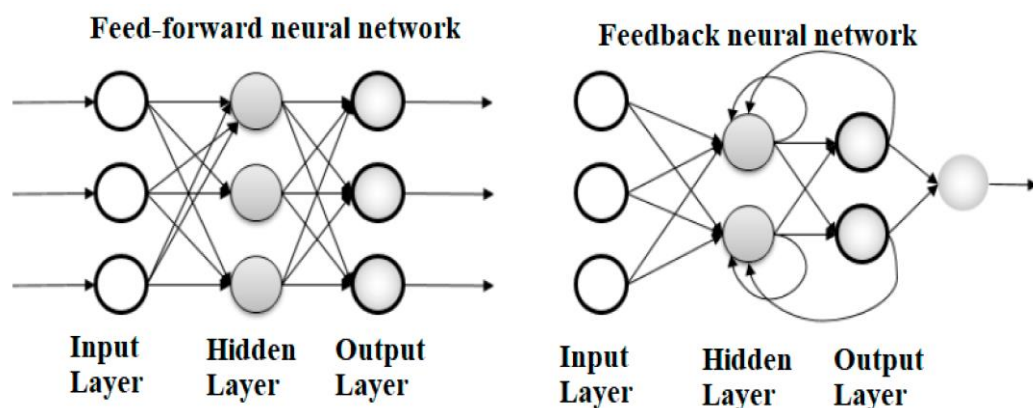


Σχήμα 3.14: Πλήρως συνδεδεμένα (αριστερά) και μερικώς συνδεδεμένα (δεξιά) δίκτυα. (Πηγή [99])



Όσον αφορά τις αρχιτεκτονικές, στις οποίες εμπίπτουν όλα τα ΤΝΔ από τα απλά δίκτυα της δεκαετίας του 1950 μέχρι και τα εξαιρετικά περίπλοκα δίκτυα της σημερινής εποχής, μπορούν να διακριθούν σε δύο βασικές κατηγορίες:

- **Δίκτυα Πρόσθιας Τροφοδότησης** ή μη κυκλικά (**Feed Forward Neural Networks**), όπου η πληροφορία ρέει «μονόδρομα», δηλαδή διαδίδεται προς μία μόνο κατεύθυνση, από την είσοδο προς την έξοδο και δεν μπορεί να επιστρέψει σε κάποιο προηγούμενο επίπεδο, εφόσον στα δίκτυα αυτά δεν υπάρχουν κύκλοι και επομένως δεν υφίσταται ανατροφοδότηση. Σε αυτήν την κατηγορία κατατάσσονται το Perceptron, τα Πολυεπίπεδα Νευρωνικά Δίκτυα (MLP), τα Συνελκτικά Νευρωνικά Δίκτυα (CNN), οι Αυτοκωδικοποιητές (Autoencoders), καθώς και τα Γεννητικά Ανταγωνιστικά Δίκτυα (GAN).
- **Δίκτυα με Ανατροφοδότηση / Αναδρομικά** ή κυκλικά (**Feedback Neural Networks**), όπου η πληροφορία ταξιδεύει «αμφίδρομα», δηλαδή μπορεί να διαδοθεί και προς τις δύο κατευθύνσεις, τόσο από την είσοδο προς την έξοδο, όσο και από την έξοδο πάλι προς την είσοδο, για αυτό και περιλαμβάνουν ένα τουλάχιστον βρόχο ανάδρασης. Κρίνεται σκόπιμο να αναφερθεί ότι τα αναδρομικά δίκτυα έχουν «μνήμη», δηλαδή απομνημονεύουν δεδομένα που έχουν προκύψει από προηγούμενες εισόδους, από τα οποία και εξαρτώνται, σε αντίθεση με τα δίκτυα πρόσθιας τροφοδότησης. Σε αυτήν την κατηγορία ανήκουν τα Αναδρομικά Νευρωνικά Δίκτυα (RNN), τα δίκτυα Hopfield, οι αυτο-οργανούμενοι χάρτες Kohonen (SOM), τα Ανταγωνιστικά Δίκτυα (Competitive Networks) και τα ART (Adaptive Resonance Theory) μοντέλα.



Σχήμα 3.15: Δίκτυα πρόσθιας τροφοδότησης (αριστερά) και αναδρομικά δίκτυα (δεξιά). (Πηγή [100])

Συμπερασματικά, προκειμένου να σχεδιαστεί σωστά ένα Τεχνητό Νευρωνικό Δίκτυο, θα πρέπει να καθοριστούν ορισμένα στοιχεία που αφορούν την αρχιτεκτονική του. Αυτά είναι:

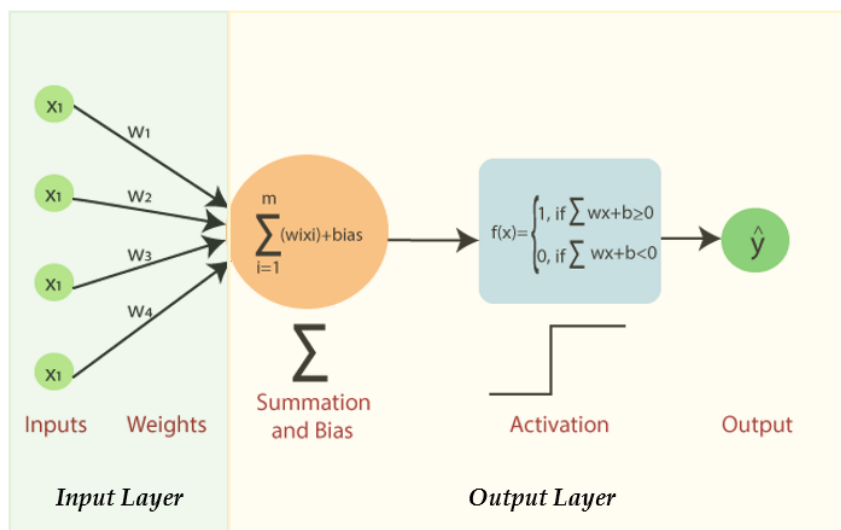
- Ο αριθμός των ενδιάμεσων κρυφών επιπέδων.
- Ο αριθμός των νευρώνων ανά επίπεδο.
- Ο τρόπος σύνδεσης των κόμβων μεταξύ τους.
- Η μορφή της συνάρτησης ενεργοποίησης, η οποία συνηθίζεται να είναι ίδια μεταξύ όλων των κρυφών επιπέδων, χωρίς, όμως, αυτό να είναι απαραίτητο.
- Οι τιμές των αρχικών βαρών μεταξύ των μονάδων, οι οποίες συνεχώς αλλάζουν κατά τη διάρκεια της εκπαίδευσης. Σε αυτές τις επαναλαμβανόμενες προσαρμογές των βαρών βασίζεται η μάθηση στα ΤΝΔ.

### 3.6.2 Perceptron

Το perceptron ή αλλιώς «αισθητήρας» αποτελεί την απλούστερη μορφή νευρωνικού δικτύου. Προτάθηκε το 1957 από τον Rosenblatt και έχει μεγάλη ιστορική σημασία, καθώς αποτέλεσε σημαντική πηγή έμπνευσης στον κλάδο της επιστήμης, ενώ ταυτοχρόνως χρησιμοποιείται μέχρι και σήμερα σαν βάση για περίπλοκα ΤΝΔ. Συχνά αναφέρεται και ως γραμμικός ταξινομητής (linear classifier), καθώς χρησιμοποιείται για την ταξινόμηση γραμμικά διαχωρίσιμων προτύπων.

Το μοντέλο αυτό αποτελείται από έναν μεμονωμένο νευρώνα που περιλαμβάνει ένα γραμμικό συνδυαστή (linear combiner), ακολουθούμενο από έναν απότομο περιοριστή (hard limiter) [75]. Στην ουσία, ως γραμμικός συνδυαστής αναφέρεται μια συνάρτηση γραμμικού συνδυασμού, τον ρόλο της οποίας παίζει ο άθροιστής των σημάτων εισόδου, τα οποία σταθμίζονται, όπως έχει αναλυθεί παραπάνω, με τα αντίστοιχα συναπτικά βάρη. Στο άθροισμα που προκύπτει, προστίθεται μια εξωτερικά εφαρμοζόμενη πόλωση και, στη συνέχεια, το αποτέλεσμα εφαρμόζεται στον απότομο περιοριστή, ο οποίος είναι, στην πραγματικότητα, μια συνάρτηση ενεργοποίησης που, όπως έχει, επίσης, προαναφερθεί, περιορίζει το πλάτος σήματος εξόδου εντός πεπερασμένων ορίων, συνήθως  $[0, 1]$  ή  $[-1, 1]$ , καθορίζοντας έτσι και το τελικό αποτέλεσμα του δικτύου.

Επομένως, πρόκειται για δίκτυο ενός επιπέδου, εφόσον περιλαμβάνει ένα μόνο επίπεδο υπολογισμού, το επίπεδο εξόδου, ενώ το επίπεδο εισόδου, συνδέεται απευθείας με αυτό, χωρίς παρεμβολή κρυφών επιπέδων και στόχος του είναι να ταξινομή τις εισόδους του σε δύο κλάσεις.



Σχήμα 3.16: Το Perceptron. (Πηγή [101])

### 3.6.3 Πολυεπίπεδα Νευρωνικά Δίκτυα

Σαν λύση στο πρόβλημα του ενός-επίπεδου perceptron, το οποίο επιτυχώς ταξινομούσε μόνο γραμμικά διαχωρίσιμα πρότυπα, εμφανίστηκαν τα Πολυεπίπεδα Νευρωνικά Δίκτυα (Multi-layer Perceptron - MLP), που χρησιμοποιούνται ως ταξινομητές ακόμη και στην περίπτωση μη γραμμικά διαχωρίσιμων προτύπων.

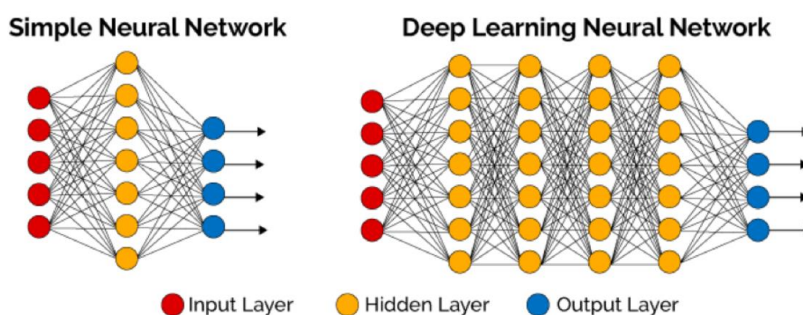
Όπως αναφέρει και το όνομά τους, πρόκειται για νευρωνικά δίκτυα, τα οποία περιλαμβάνουν τουλάχιστον τρία επίπεδα: ένα επίπεδο εισόδου με πηγαίους νευρώνες, τουλάχιστον ένα κρυφό επίπεδο με υπολογιστικούς νευρώνες και ένα επίπεδο εξόδου με νευρώνες που, επίσης, εκτελούν υπολογισμούς [94]. Πρόκειται για δίκτυα πλήρως συνδεδεμένα (fully connected), με αρχιτεκτονική πρόσθιας τροφοδότησης. Αυτό πρακτικά σημαίνει ότι όλοι οι νευρώνες ενός επιπέδου συνδέονται με όλους τους νευρώνες του αμέσως προηγούμενου και ότι τα σήματα εισόδου διαδίδονται επίπεδο προς επίπεδο, ξεκινώντας από το στρώμα εισόδου, με κατεύθυνση προς τα εμπρός.

Όσον αφορά τα κρυφά επίπεδα των MLP, έχουν ονομαστεί κατ' αυτόν τον τρόπο, λόγω του ότι αποκρύπτουν την επιθυμητή έξοδό τους, η οποία τελικά καθορίζεται από το εκάστοτε επίπεδο [94]. Ωστόσο, διαδραματίζουν σημαντικό ρόλο στη διαδικασία εκπαίδευσης, καθώς οι κρυφοί νευρώνες, μέσω των μη γραμμικών μετασχηματισμών που εφαρμόζουν στα σήματα εισόδου, χρησιμοποιούνται για την εξαγωγή χρήσιμης πληροφορίας. Πράγματι, είναι υπεύθυνοι για την ανίχνευση χαρακτηριστικών (feature detection) των δεδομένων εκπαίδευσης κατά τη διαδικασία μάθησης και για αυτόν το λόγο,

θεωρούνται συστατικό στοιχείο των νευρωνικών δικτύων, δίνοντάς τους τη δυνατότητα να εκπαιδεύονται σε περίπλοκα προβλήματα και να πετυχαίνουν υψηλή απόδοση. Όταν ένα ΤΝΔ περιλαμβάνει περισσότερα του ενός κρυφά στρώματα, τότε θεωρείται Βαθύ Νευρωνικό Δίκτυο. Σε αυτήν την περίπτωση, καθένα κρυφό επίπεδο είναι υπεύθυνο για την εξαγωγή διαφορετικού τύπου χαρακτηριστικών, βοηθώντας, έτσι, στην ανακάλυψη των κρυμμένων μοτίβων και καταλήγοντας σε αποτελέσματα μεγαλύτερης ακρίβειας.

### 3.6.4 Βαθιά Νευρωνικά Δίκτυα

Τα Βαθιά Νευρωνικά Δίκτυα (Deep Neural Networks - DNN) ανήκουν στην ευρύτερη κατηγορία των ΤΝΔ, με την ιδιαιτερότητα ότι απαρτίζονται από δύο ή περισσότερα κρυφά επίπεδα. Ο χαρακτηρισμός «Βαθιά», ωστόσο, δεν αναφέρεται μόνο στον αριθμό των επιπέδων, αλλά και στον μεγαλύτερο αριθμό νευρώνων σε κάθε στρώμα. Αυτό έχει σαν αποτέλεσμα, όπως είναι λογικό, την αύξηση της πολυπλοκότητας του δικτύου και κατ'επέκταση την ικανότητά του να αναγνωρίζει πρότυπα υψηλότερου επιπέδου, δηλαδή μοτίβα που δύσκολα θα εντόπιζε ένα απλό Νευρωνικό Δίκτυο. Έτσι, τα Βαθιά Νευρωνικά Δίκτυα γνώρισαν μεγάλη δημοσιότητα στον χώρο της Μηχανικής Μάθησης, λόγω της σημαντικά αυξημένης απόδοσής τους, κατά κόρον, σε προβλήματα όρασης υπολογιστών, επεξεργασίας φωνής και γλώσσας. Ωστόσο, συχνά αποτελεί πρόκληση ο σχεδιασμός τους, διότι δεν είναι πάντοτε εύκολο να βρεθεί η χρυσή τομή μεταξύ απόδοσης των εν λόγω δικτύων και δαπάνης υπολογιστικών πόρων.



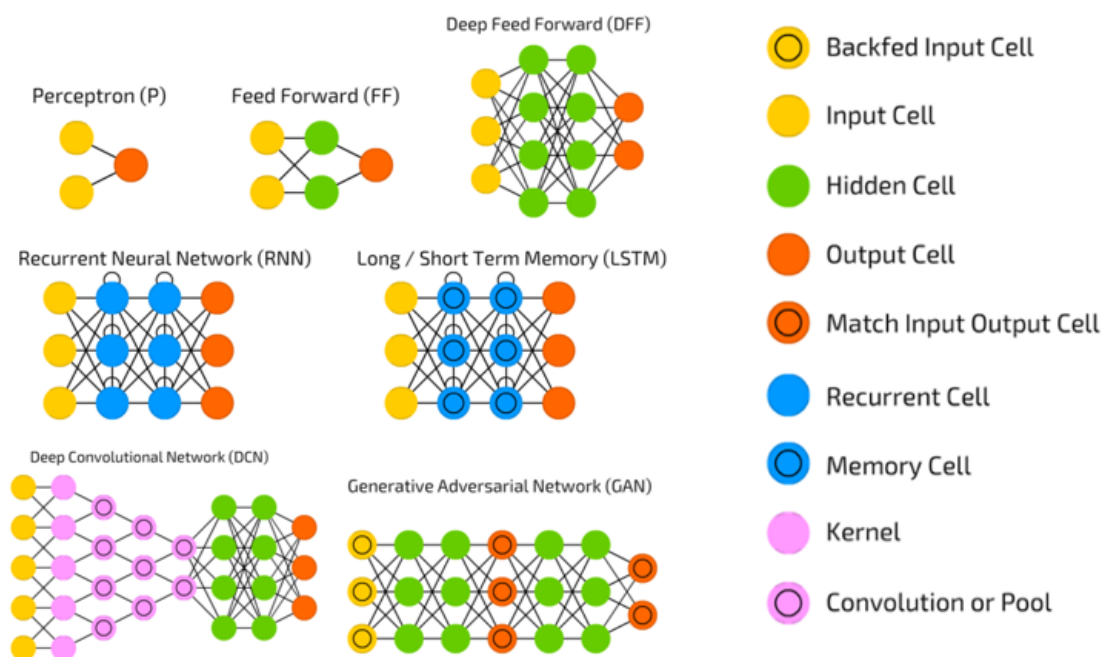
Σχήμα 3.17: Απλό vs Βαθύ Νευρωνικό Δίκτυο. (Πηγή [102])

Τα βαθιά νευρωνικά δίκτυα, λειτουργούν όπως και τα απλά ΤΝΔ, δηλαδή μαθαίνουν «νέες δεξιότητες» κατά τη φάση της εκπαίδευσης (training phase) από τα δεδομένα που τους παρέχονται και, στη συνέχεια, η αποκτηθείσα γνώση εφαρμόζεται σε άγνωστα δεδομένα στη φάση των προβλέψεων (prediction phase). Η διαφορά τους, η οποία σχετίζεται με την ύπαρξη μεγάλου αριθμού ενδιάμεσων επιπέδων μεταξύ εισόδου και

εξόδου, έγκειται στο γεγονός ότι κάθε κρυφό στρώμα υπηρετεί διαφορετικό σκοπό, όσον αφορά την εξαγωγή χρήσιμης πληροφορίας και, με αυτόν τον τρόπο το δίκτυο οδηγείται σταδιακά στο τελικό αποτέλεσμα πετυχαίνοντας υψηλότερη απόδοση. Πιο συγκεκριμένα, τα πρώτα κρυφά επίπεδα, που βρίσκονται πλησιέστερα στο input layer, εξάγουν χαμηλού επιπέδου χαρακτηριστικά, που αφορούν στην ανίχνευση ακμών, γραμμών ή περιγραμμάτων, ενώ όσο κατευθυνόμαστε προς το output layer, ανιχνεύουν μοτίβα πιο σύνθετα, πιο «ολοκληρωμένα», τα οποία πλησιάζουν περισσότερο προς την τελική απόκριση του συστήματος, κάτι λογικό, αν αναλογιστούμε ότι έχουν λάβει όλη την πληροφορία που προκύπτει από τα προηγούμενα κρυφά επίπεδα.

Τα πιο συχνά χρησιμοποιούμενα Βαθιά Νευρωνικά Δίκτυα είναι τα εξής:

- Συνελκτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks - CNN)
- Αναδρομικά Νευρωνικά Δίκτυα (Recurrent Neural Networks - RNN)
- Δίκτυα Μακράς Βραχύχρονης Μνήμης (Long Short Term Memory Networks - LSTM)
- Γεννητικά Ανταγωνιστικά Δίκτυα (Generative Adversarial Networks - GAN)

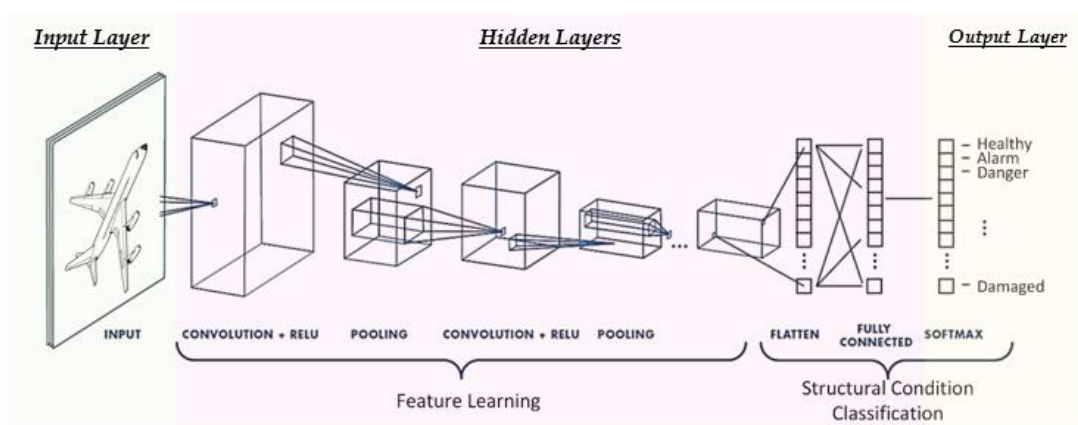


Σχήμα 3.18: Συγκριτική παρουσίαση Νευρωνικών Δικτύων. (Πηγή [103])

### 3.7 Συνελικτικά Νευρωνικά Δίκτυα

Τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks - CNN) ανήκουν στην ευρύτερη κατηγορία των Πολυεπίπεδων ΤΝΔ με πρόσθια τροφοδότηση, κατέχοντας μία από τις πρώτες θέσεις στις πιο συχνά χρησιμοποιούμενες αρχιτεκτονικές δικτύων Βαθιάς Μάθησης. Η συμβολή τους είναι καθοριστική όσον αφορά στην ανάλυση και ταξινόμηση εικόνων και βίντεο, καθώς εξειδικεύονται στην επεξεργασία δεδομένων με γνωστή πλεγματοειδή τοπολογία (grid-like topology), τόσο μονοδιάστατων (1D grid), όπως είναι οι χρονοσειρές, όσο και δισδιάστατων (2D grid pixels) όπως είναι οι εικόνες με τα εικονοστοιχεία τους να σχηματίζουν πλέγματα δύο διαστάσεων, αλλά ακόμη και τρισδιάστατων (3D grid pixels) όπως τα βίντεο [104]. Παράλληλα, όμως, με επιτυχία χρησιμοποιούνται και σε άλλα προβλήματα Μηχανικής Μάθησης, όπως είναι τα συστήματα συστάσεων και η επεξεργασία φυσικής γλώσσας.

Ως Πολυεπίπεδα Βαθιά Νευρωνικά Δίκτυα, τα CNN απαρτίζονται από ένα επίπεδο εισόδου, από κρυφά επίπεδα και ένα επίπεδο εξόδου. Ωστόσο, όπως προδίδει και το όνομά τους, η ιδιαιτερότητα των εν λόγω δικτύων συνίσταται στο ότι περιλαμβάνουν συνελικτικά στρώματα, δηλαδή στρώματα, τα οποία βασίζονται στη συνέλιξη, μια μαθηματική πράξη που μετρά την «επικάλυψη» μεταξύ δύο συναρτήσεων, όταν η μία έχει υποστεί ανάκλαση και μετατόπιση. Έτσι, τα δίκτυα αυτά, στη θέση του απλού πολλαπλασιασμού των εκάστοτε εισόδων με τα αντίστοιχα συναπτικά βάρη, εκτελούν τη γραμμική πράξη της συνέλιξης μεταξύ της δισδιάστατης εισόδου και ενός δισδιάστατου πίνακα βαρών που ονομάζεται πυρήνας (kernel) ή φίλτρο, σε ένα τουλάχιστον από τα επίπεδά τους.



Σχήμα 3.19: Συνελικτικό Νευρωνικό Δίκτυο. (Πηγή [105])

Το κύριο πλεονέκτημά τους έγκειται στο γεγονός ότι αυτόματα ανιχνεύουν τα σημαντικά χαρακτηριστικά από ανεπεξέργαστα δεδομένα, χωρίς, δηλαδή, να είναι

απαραίτητη η ανθρώπινη παρέμβαση, εν συγκρίσει με τα ΤΝΔ που προϋπήρχαν και τα οποία απαιτούσαν στο επίπεδο εισόδου να τους παρέχεται έτοιμη η χρήσιμη πληροφορία από τα δεδομένα εκπαίδευσης σε συγκεκριμένη μορφή. Μια ακόμη διαφορά τους εντοπίζεται στον αριθμό των παραμέτρων, ο οποίος εμφανίζεται σημαντικά μεγαλύτερος στην περίπτωση των MLP, κάτι που δίνει, επίσης, προβάδισμα στα συνελκτικά δίκτυα, εφόσον μειώνονται οι απαιτήσεις σε υπολογιστική ισχύ. Όλα αυτά, σε συνδυασμό με την ευκολία που παρουσιάζουν οι συνελκίξεις στο να παραλληλιζονται μεταξύ των πυρήνων GPU [106], ανέδειξαν την υπολογιστική αποτελεσματικότητα των CNN και οδήγησαν στην ευρεία χρήση τους από τον κόσμο της επιστημονικής κοινότητας.

Ιστορικά, πηγή έμπνευσης των ConvNets, όπως συχνά ονομάζονται τα συνελκτικά δίκτυα, υπήρξε μια σειρά ερευνών που διεξήχθησαν το 1962 από τους David Hubel και Torsten Wiesel [107], [108], [109], αναφορικά με τους «απλούς» και «σύνθετους» βιολογικούς νευρώνες του οπτικού φλοιού ανώτερων θηλαστικών, προτείνοντας ότι και τα δύο είδη κυττάρων χρησιμοποιούνται στην αναγνώριση προτύπων. Λίγο αργότερα, το 1980, ο Kunihiko Fukushima στηριζόμενος στην έρευνα των Hubel & Wiesel, πρότεινε το μοντέλο Neocognitron, ένα νευρωνικό δίκτυο που εκπαιδεύεται χωρίς επίβλεψη και είναι ικανό να αναγνωρίζει μοτίβα βάσει της γεωμετρικής ομοιότητας των σχημάτων τους, χωρίς να επηρεάζονται από τη θέση τους στο χώρο [110]. Στην εργασία αυτή βασίστηκαν οι Yann LeCun κ.α., οι οποίοι προσεγγίζουν πλέον τα μοντέρνα CNN, προτείνοντας, στα τέλη της δεκαετίας του 1990, ότι η αναγνώριση προτύπων μπορεί να πετύχει καλύτερα αποτελέσματα, μέσω της αυτόματης μάθησης, συγκρινόμενη με παραδοσιακές μεθόδους αναγνώρισης [111]. Στη δημοσίευσή τους εξήγησαν τις αρχιτεκτονικές ιδέες για το σχεδιασμό ενός τέτοιου συνελκτικού δικτύου, «αναλλοίωτου» ως προς τη μετατόπιση, την κλιμάκωση ή τη στρέβλωση, δηλαδή ενός δικτύου, του οποίου το τελικό αποτέλεσμα δεν επηρεάζεται από αυτά ή άλλα είδη παραμορφώσεων της εικόνας - εισόδου. Ονομάστηκε LeNet-5.

Μέσα στα επόμενα χρόνια, οι έρευνες συνεχίστηκαν και τα συνελκτικά μοντέλα συνεχώς εξελίσσονταν. Ωστόσο, ένα από τα πρώτα σύγχρονα CNN που αναπτύχθηκαν και κίνησε το ενδιαφέρον της επιστημονικής κοινότητας, λόγω του ότι πέτυχε υψηλές επιδόσεις σε ένα απαιτητικό πρόβλημα ταξινόμησης υψηλής ανάλυσης εικόνων, υπερνικώντας τις συμβατικές μεθόδους όρασης υπολογιστών, ήταν το AlexNet που σχεδιάστηκε από τους Alex Krizhevsky κ.α., το 2012 [112]. Έκτοτε και μέχρι σήμερα, έχουν αναπτυχθεί πολλά μοντέλα Συνελκτικών Νευρωνικών Δικτύων που είναι ακόμη πιο «Βαθιά» από το AlexNet και χρησιμοποιούνται ευρέως σε μια πληθώρα εφαρμογών, ακόμη πιο «μοντέρνων», όπως η παρακολούθηση αντικειμένων σε βίντεο (video tracking),



η κατάτμηση εικόνας (image segmentation) και η αναγνώριση αντικειμένων (object detection).

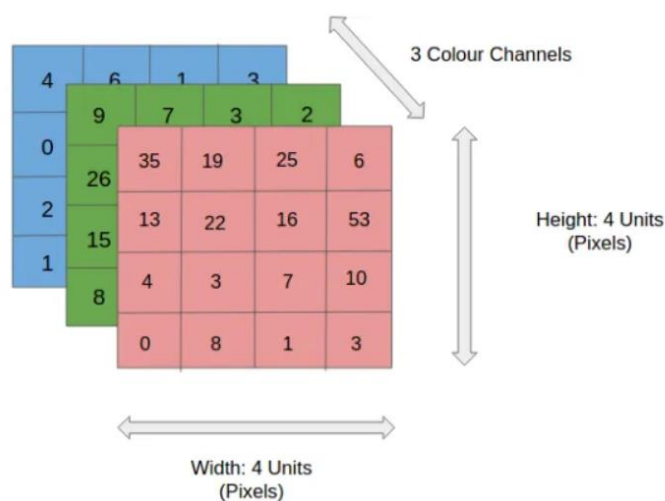
### 3.7.1 Επίπεδα επεξεργασίας

Αυτό που κάνει ιδιαίτερα τα Συνελκτικά Νευρωνικά Δίκτυα είναι οι πρωτότυπες αρχιτεκτονικές ιδέες πίσω από τον σχεδιασμό τους. Τα κρυφά επίπεδα, από τα οποία απαρτίζονται δεν είναι όλα πανομοιότυπα μεταξύ τους, αλλά υπάρχει μια δομημένη ακολουθία στρωμάτων, όπου το καθένα επιτελεί από μια διαφορετική λειτουργία, συνδράμοντας όλα στην επίτευξη μιας «καλής» πρόβλεψης. Να σημειωθεί ότι αυτή τους η αρχιτεκτονική, δεν είναι μόνο καλή στην εκμάθηση χαρακτηριστικών, αλλά και επεκτάσιμη σε τεράστια σύνολα δεδομένων. Σε όλα τα CNN συναντώνται τα επίπεδα επεξεργασίας που παρουσιάζονται παρακάτω με τον αριθμό τους να ποικίλλει, καθορίζοντας, έτσι, το βάθος του εκάστοτε δικτύου.

#### Επίπεδο Εισόδου

Όπως σε όλα τα ΤΝΔ, αποτελεί το πρώτο στρώμα του δικτύου, μέσω του οποίου εισέρχεται η πληροφορία από το περιβάλλον, με σκοπό την περαιτέρω επεξεργασία της από τα στρώματα που ακολουθούν. Στα Συνελκτικά Δίκτυα, σαν είσοδος μπορεί να χρησιμοποιηθεί μία ή περισσότερες εικόνες, η διάσταση των οποίων καθορίζει και τη «διάσταση» του επιπέδου αυτού.

Σε αυτό το σημείο, να εξηγήσουμε ότι μια εικόνα είναι, στην ουσία, ένας πίνακας τιμών, με κάθε «κουτάκι» του να αντιπροσωπεύει την τιμή ενός εικονοστοιχείου της. Για να οριστεί χρειάζονται τρεις παράμετροι, το μήκος, το πλάτος και ο αριθμός καναλιών της.



Σχήμα 3.20: Παράδειγμα μίας RGB εικόνας διαστάσεων (4×4×3). (Πηγή [113])



Ο τελευταίος εξαρτάται από τον αριθμό των τιμών που απαιτούνται για να αναπαριστήσουν ένα εικονοστοιχείο της. Πιο αναλυτικά, εάν είναι ασπρόμαυρη κάθε pixel αντιστοιχίζεται με 1 τιμή, οπότε η εικόνα έχει 1 κανάλι, ενώ εάν είναι έγχρωμη κάθε pixel αντιστοιχίζεται συνήθως με 3 τιμές, οπότε έχει 3 κανάλια. Βάσει αυτών, το επίπεδο εισόδου ενός CNN, δέχεται δεδομένα διαστάσεων:

$$(\text{Πλήθος Εικόνων}) \times (\text{Μήκος Εικόνων}) \times (\text{Πλάτος Εικόνων}) \times (\text{Αριθμός Καναλιών})$$

### Συνελικτικό Επίπεδο

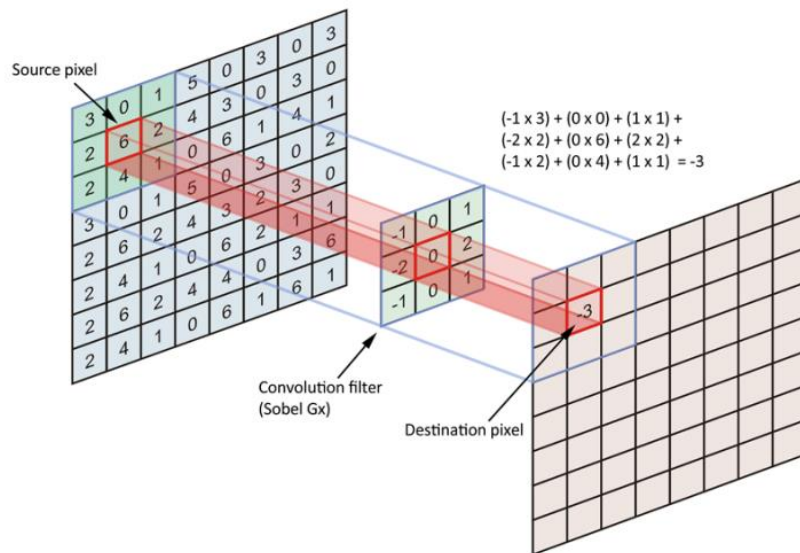
Όπως φανερώνει και το όνομά του, αποτελεί το κύριο δομικό συστατικό των Συνελικτικών Νευρωνικών Δικτύων, τα οποία μπορεί να περιλαμβάνουν ένα ή περισσότερα τέτοια συνελικτικά στρώματα (convolutional layers). Είναι το επίπεδο που σηκώνει το βαρύτερο υπολογιστικό φορτίο, εφόσον σε αυτό εκτελείται η πράξη της συνέλιξης και είναι υπεύθυνο για την εξαγωγή της χρήσιμης πληροφορίας. Αυτό επιτυγχάνεται μέσω διαφόρων μετασχηματισμών που εφαρμόζει στα δεδομένα που δέχεται ως είσοδο. Σαν αποτέλεσμα αποδίδει πίνακες χαρακτηριστικών, οι οποίοι ανάλογα με το βάθος στο οποίο βρίσκεται το εκάστοτε συνελικτικό στρώμα, απαρτίζονται, όπως έχει προαναφερθεί, από χαμηλού έως και υψηλού επιπέδου χαρακτηριστικά.

Πιο συγκεκριμένα, στα συνελικτικά επίπεδα τα διανύσματα εισόδου συνελίσσονται με πυρήνες (kernels), που συχνά ονομάζονται και φίλτρα (filters), οι οποίοι είναι, ουσιαστικά, πίνακες συναπτικών βαρών. Σε μια δισδιάστατη εικόνα  $I(x,y)$  η πράξη της συνέλιξης με έναν, επίσης, δισδιάστατο πυρήνα  $K(x,y)$  περιγράφεται μαθηματικά ως εξής:

$$S_{i,j} = (I * K)(i,j) = \sum_m \sum_n I(m,n)K(i-m,j-n)$$

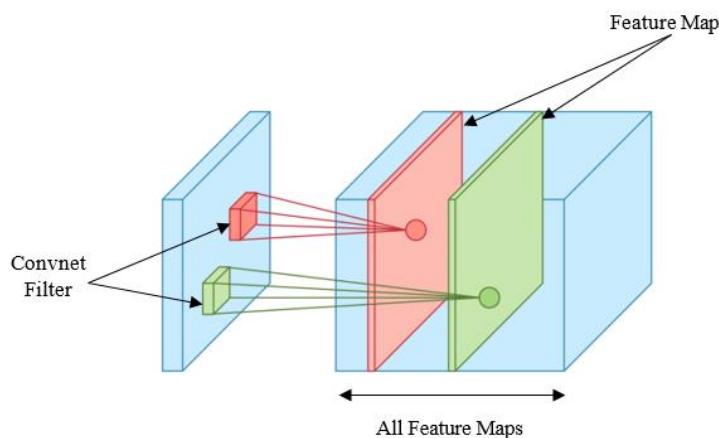
Η πράξη της συνέλιξης χρησιμοποιείται για τη «συγχώνευση» δύο συνόλων πληροφοριών, στην προκειμένη περίπτωση, της εικόνας-εισόδου με τον πυρήνα βαρών. Στην πραγματικότητα, εκτελείται ένα προς ένα πολλαπλασιασμός μεταξύ των στοιχείων των δύο πινάκων (εικόνας και πυρήνα) και ακολουθεί πρόσθεση των αποτελεσμάτων, με την τελική τιμή να αποθηκεύεται σε έναν τρίτο πίνακα. Ωστόσο, συνήθως, τα δεδομένα εισόδου με το φίλτρο δεν έχουν τις ίδιες διαστάσεις, οπότε η διαδικασία έχει ως εξής: το φίλτρο ολισθαίνει πάνω στην εικόνα-πίνακα, μετακινούμενο κατά ένα καθορισμένο αριθμό στοιχείων κάθε φορά, το βήμα ολίσθησης (stride) (που συχνά λαμβάνει την τιμή 1, αλλά εξαρτάται από την επικάλυψη που επιθυμούμε για τις «χωρικές γειτονιές» της εικόνας), έως ότου καλύψει ολόκληρη την επιφάνειά της. Η συνέλιξη, σε κάθε βήμα αυτής της

διαδικασίας, εκτελείται μεταξύ του εν λόγω φίλτρου και του τμήματος της εισόδου που καλύπτει το φίλτρο αυτό (τοπικό δεκτικό πεδίο). Τα αποτελέσματα που προκύπτουν από κάθε συνέλιξη τοποθετούνται στον σημαντικό εκείνο πίνακα, που ονομάζεται χάρτης χαρακτηριστικών (feature map) ή χάρτης ενεργοποίησης (activation map). Έτσι, για κάθε πυρήνα που συνελίσσεται με μία εικόνα δημιουργείται ένας μοναδικός χάρτης χαρακτηριστικών.



Σχήμα 3.21: Συνελικτικό Επίπεδο. (Πηγή [114])

Ωστόσο, μία εικόνα δεν πολλαπλασιάζεται, συνήθως, με ένα και μοναδικό φίλτρο, αλλά με παραπάνω, καθένα από τα οποία εξάγει και διαφορετικά σημεία ενδιαφέροντος. Οπότε γίνεται κατανοητό, ότι προκύπτουν περισσότερα του ενός feature maps. Στη συνέχεια, οι χάρτες χαρακτηριστικών στοιβάζονται, δομώντας τελικά έναν πίνακα χαρακτηριστικών, ο οποίος αποτελεί και την έξοδο κάθε συνελικτικού επιπέδου.



Σχήμα 3.22: Οι χάρτες χαρακτηριστικών στοιβάζονται και δημιουργούν έναν πίνακα χαρακτηριστικών.

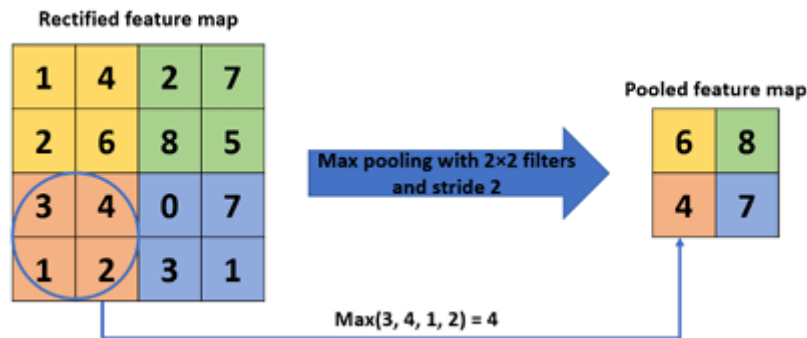
(Πηγή [115])

Ο πίνακας χαρακτηριστικών είναι τρισδιάστατος με το βάθος του να καθορίζεται από το πλήθος των χαρτών χαρακτηριστικών, δηλαδή από το πλήθος των πυρήνων που χρησιμοποιήθηκαν στο εκάστοτε επίπεδο. Όσον αφορά το πλάτος και το ύψος του, αυτά καθορίζονται από τις διαστάσεις των δεδομένων εισόδου, το μέγεθος του πυρήνα, το βήμα ολίσθησης (stride), το οποίο όσο μεγαλώνει, τόσο μικρότεροι χάρτες προκύπτουν, καθώς επίσης και την επέκταση των περιθωρίων (padding) της αρχικής εισόδου, η οποία συχνά προστίθεται για τη διατήρηση του μεγέθους των χαρτών ενεργοποίησης, διαφορετικά θα συρρικνώνονταν σε κάθε επίπεδο, πράγμα που δεν είναι επιθυμητό.

Τέλος, κρίνεται σημαντικό να αναφερθεί ότι, όπως όλα τα ΤΝΔ, έτσι και τα Συνελικτικά Δίκτυα χαρακτηρίζονται από μη-γραμμικότητα. Αυτό καθίσταται εφικτό στα σύγχρονα CNN, περνώντας το αποτέλεσμα κάθε συνέλιξης μέσω μιας συνάρτησης ενεργοποίησης και πιο συγκεκριμένα της ReLU. Έτσι, οι τιμές που τελικά αποθηκεύονται στους χάρτες χαρακτηριστικών δεν είναι, στην πραγματικότητα, τα αθροίσματα των συνέλιξεων, όπως εξηγήθηκε παραπάνω για λόγους απλότητας, αλλά τα αποτελέσματα που προκύπτουν όταν η συνάρτηση ReLU εφαρμόζεται σε αυτά.

### Συγκεντρωτικό Επίπεδο

Τα συγκεντρωτικά επίπεδα (pooling layers), συνήθως, έπονται των συνελικτικών, και έτσι ως είσοδο, λαμβάνουν τρισδιάστατους πίνακες χαρακτηριστικών. Σκοπός τους είναι να κάνουν την αναπαράσταση σχετικά αμετάβλητη σε μικρές μετατοπίσεις, αλλά και να ενισχύσουν την ικανότητα γενίκευσης που έχει το δίκτυο, δίνοντας έμφαση στην παρουσία χαρακτηριστικών στο διάνυμα εισόδου και όχι τόσο στον εντοπισμό της θέσης τους [104]. Λειτουργούν παρόμοια με τα συνελικτικά στρώματα, αν και δεν έχουν παραμέτρους, όπως ήταν τα συναπτικά βάρη των πυρήνων. Πράγματι, αποτελούνται από ένα παράθυρο σταθερού μεγέθους που ολισθαίνει, σε όλη την επιφάνεια του κάθε χάρτη χαρακτηριστικών, μετακινούμενο κάθε φορά κατά ένα καθορισμένο βήμα ολίσθησης και με αυτόν τον τρόπο ομαδοποιούν τις γειτονικές τιμές του χάρτη. Στη συνέχεια, στα γειτονικά αυτά στοιχεία εφαρμόζουν κάποιου είδους υπολογισμό, καταλήγοντας σε μια τελική τιμή για κάθε γειτονιά, η οποία αποθηκεύεται σε ένα νέο χάρτη χαρακτηριστικών. Συνήθως, η μέθοδος υπολογισμού που επιλέγεται ονομάζεται max pooling και στην ουσία, επιλέγει το μέγιστο από το στοιχεία της «γειτονιάς» του χάρτη. (Παλαιότερα χρησιμοποιούνταν και το average pooling, που υπολόγιζε το μέσο όρο των στοιχείων της εκάστοτε «γειτονιάς», όμως πλέον δεν προτιμάται.) Τόσο το μέγεθος του παραθύρου, όσο και το βήμα ολίσθησης, καθορίζουν τις διαστάσεις του τελικού αποτελέσματος.



Σχήμα 3.23: Συγκεντρωτικό επίπεδο με χρήση max pooling. (Πηγή [116])

Βάσει της διαδικασίας αυτής, οι διαστάσεις του αρχικού πίνακα χαρακτηριστικών μειώνονται σημαντικά, γι' αυτό και το επίπεδο αυτό αναφέρεται συχνά και σαν επίπεδο υποδειγματοληψίας. Πιο συγκεκριμένα, οι διαστάσεις που υφίστανται μείωση είναι το πλάτος και το ύψος και όχι το βάθος του, που σχετίζεται με το πλήθος των χαρτών ενεργοποίησης. Αυτό συμβαίνει, καθώς το ολισθαίνον παράθυρο από μια «γειτονιά» στοιχείων στον αρχικό πίνακα χαρακτηριστικών καταλήγει σε ένα μόνο στοιχείο στον τελικό πίνακα χαρακτηριστικών, οπότε μειώνεται ο αριθμός των στοιχείων του κάθε χάρτη και κατ' επέκταση, το πλάτος και το ύψος του. Όμως, η ίδια διαδικασία εκτελείται παράλληλα σε κάθε χάρτη χαρακτηριστικών, με αποτέλεσμα το πλήθος τους να μη μεταβάλλεται.

Αυτή η μείωση των διαστάσεων που πραγματοποιείται στα συγκεντρωτικά στρώματα έχει σαν αποτέλεσμα την ελάττωση των παραμέτρων του συστήματος. Σαν φυσικό επακόλουθο έρχεται η μείωση του χρόνου εκπαίδευσης και της υπολογιστικής ισχύος, που δίνει σημαντικό πλεονέκτημα στα δίκτυα αυτά. Επίσης, αντιμετωπίζεται το πρόβλημα της υπερπροσαρμογής (overfitting), δεδομένου του ότι με την ελάττωση των παραμέτρων, παραλείπονται ασήμαντες λεπτομέρειες και θόρυβος των δεδομένων εκπαίδευσης που συχνά οδηγούν σε χαμηλότερη απόδοση του μοντέλου.

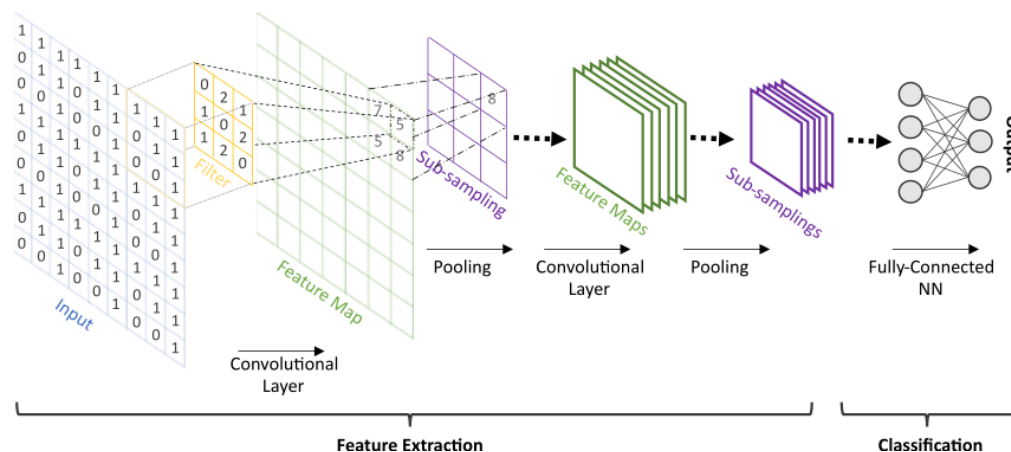
### Πλήρως Συνδεδεμένο Επίπεδο

Τα πλήρως συνδεδεμένα επίπεδα (Fully connected layers) αποτελούν τα τελευταία στρώματα των συνελκτικών δικτύων. Συναντώνται μετά από μια σειρά συνελκτικών και συγκεντρωτικών επιπέδων και οργανώνονται σε ομάδες ενός ή περισσότερων πλήρως συνδεδεμένων στρωμάτων. Όπως φαίνεται και από το όνομά τους, κάθε νευρώνας του επιπέδου αυτού συνδέεται με όλους τους νευρώνες του επόμενου, οπότε στην πραγματικότητα οι ομάδες αυτές συγκροτούν ένα πολυεπίπεδο νευρωνικό δίκτυο. Ως

είσοδο λαμβάνουν τον τρισδιάστατο πίνακα χαρακτηριστικών που προέκυψε σαν έξοδος από το αμέσως προηγούμενο συγκεντρωτικό επίπεδο. Ωστόσο, όπως έχει προαναφερθεί, αυτού του είδους τα δίκτυα αναμένουν κάποια μονοδιάστατη είσοδο. Επομένως, εφαρμόζεται η μέθοδος της ισοπέδωσης (flatten) στα δεδομένα εισόδου του επιπέδου αυτού, κατά την οποία ο τρισδιάστατος πίνακας μετασχηματίζεται σε μονοδιάστατο διάνυσμα, χωρίς, φυσικά, να έχουμε απώλεια πληροφορίας.

Τα πλήρως συνδεδεμένα επίπεδα αποσκοπούν στην αξιοποίηση των σημείων ενδιαφέροντος που εξήχθησαν από τα προηγούμενα συνελκτικά μπλοκ, ώστε να καταλήξουν στο τελικό αποτέλεσμα. Ο ρόλος τους έγκειται στην ικανότητα εκμάθησης μη γραμμικών συνδυασμών των σημαντικών χαρακτηριστικών της εισόδου, στην εξαγωγή προτύπων και κρυμμένων μοτίβων. Γι' αυτό και συχνά τα συνελκτικά δίκτυα χρησιμοποιούνται σε εφαρμογές ταξινόμησης εικόνων, όπου το τελευταίο πλήρως συνδεδεμένο επίπεδό τους, βάσει της χρήσιμης πληροφορίας, αναδεικνύει την κλάση της εκάστοτε εισόδου.

Βάσει όλων των παραπάνω ένα ConvNet θα μπορούσε να αποδοθεί σχηματικά ως εξής:



Σχήμα 3.24: Σχηματική αναπαράσταση των επιπέδων επεξεργασίας ενός CNN. (Πηγή [117])

### 3.8 Μεταφορά Μάθησης

Η Μεταφορά Μάθησης (Transfer Learning) είναι μια μέθοδος Μηχανικής Μάθησης, κατά την οποία ένα μοντέλο που αναπτύχθηκε για κάποια εργασία επαναχρησιμοποιείται, εκπαιδευμένο πλέον, για κάποια άλλη, συνήθως παρεμφερή. Στην ουσία, αυτό που συμβαίνει είναι ότι η προϋπάρχουσα γνώση «μεταφέρεται», με σκοπό την επίλυση ενός νέου προβλήματος. Πρόκειται για μια δημοφιλή προσέγγιση στο πεδίο του Deep Learning, όπου τα προ-εκπαιδευμένα μοντέλα συχνά χρησιμοποιούνται σε εργασίες όρασης

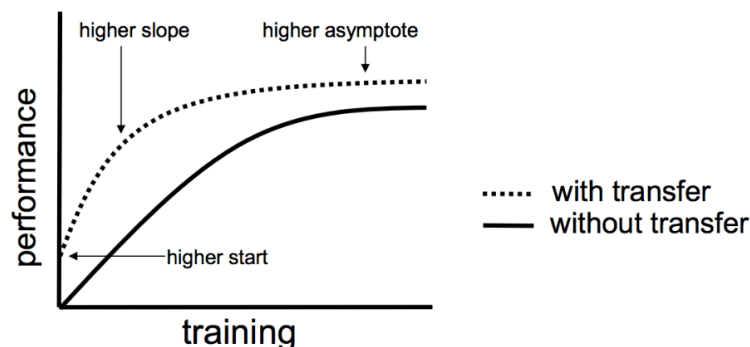
υπολογιστών και επεξεργασίας φυσικής γλώσσας, δεδομένων των τεράστιων υπολογιστικών πόρων που απαιτούνται για την ανάπτυξη και εκπαίδευση μοντέλων νευρωνικών δικτύων σε αυτά τα προβλήματα. Αυτό έχει σαν αποτέλεσμα την εξοικονόμηση χρόνου, χώρου, αλλά και κόστους, φυσικά, ενώ παράλληλα παρατηρείται και σημαντική βελτίωση της απόδοσης στο εκάστοτε πρόβλημα. Βέβαια, η μεταφορά μάθησης δεν αποτελεί αποκλειστικά τομέα μελέτης για τη βαθιά μάθηση, αλλά συχνά σχετίζεται με προβλήματα όπως η μάθηση πολλαπλών εργασιών (multi-task learning) και η εννοιολογική απόκλιση (concept drift).

Ένας ορισμός που έχει δοθεί για τη μεταφορά μάθησης σε μία από τις πρώτες ολοκληρωμένες έρευνες [118] πάνω σε αυτό τον τομέα, υποστηρίζει ότι δοθέντος ενός τομέα-πηγή  $D_S$  (source domain) και μιας διεργασίας εκμάθησης-πηγή  $T_S$  (source task), καθώς και ενός τομέα-στόχος  $D_T$  (target domain) και αντίστοιχα διεργασίας εκμάθησης-στόχος  $T_T$  (target task), σκοπός της μεταφοράς μάθησης αποτελεί η βελτίωση της εκμάθησης της συνάρτησης πρόβλεψης  $f_T(\cdot)$  για τον τομέα-στόχο, χρησιμοποιώντας τη γνώση από τα  $D_S$  και  $T_S$ , όπου  $D_S \neq D_T$  και  $T_S \neq T_T$ . Με άλλα λόγια, η μεταφορά μάθησης επιτρέπει την αξιοποίηση της γνώσης από κάποιον τομέα - πηγή (source domain), με σκοπό τη βελτίωση της απόδοσης του εκάστοτε μοντέλου στον τομέα - στόχο (target domain).

Σε πρακτικό επίπεδο, η μεταφορά μάθησης συμβαίνει ως εξής, αφαιρείται από το εν λόγω προ-εκπαιδευμένο μοντέλο το επίπεδο, όπου γίνεται η ανάθεση κλάσης, δηλαδή το τελευταίο επίπεδο του δικτύου ή συχνά και προγενέστερα αυτού. Στη συνέχεια, ανάλογα με την εφαρμογή για την οποία πρόκειται να χρησιμοποιηθεί το μοντέλο, υπάρχουν δύο επιλογές. Στην πρώτη περίπτωση, τα τελευταία αυτά στρώματα αντικαθίστανται με νέα, κατάλληλα να επιλύσουν το πρόβλημα-στόχος. Έπειτα, το μοντέλο επανεκπαιδεύεται με χρήση των καινούριων δεδομένων και αυτό έχει σαν αποτέλεσμα να συνεχίσουν να προκαλούνται μικρές τροποποιήσεις των βαρών σε όλα τα επίπεδα του μοντέλου, όχι μόνο στα προσφάτως προστιθέμενα επίπεδα. Στη δεύτερη περίπτωση, δεν προστίθεται νέα layers στη θέση αυτών που αφαιρέθηκαν, αλλά το συγκεκριμένο μοντέλο διατηρείται ως έχει, καθώς στόχος είναι να χρησιμοποιηθεί για την εξαγωγή χρήσιμων χαρακτηριστικών, όπως και γίνεται στα πλαίσια της παρούσας εργασίας, και έτσι δεν είναι απαραίτητα αυτά τα τελικά στρώματα.

Έχει αποδειχθεί ότι η μεταφορά μάθησης έχει ορισμένα οφέλη όσον αφορά τη διαδικασία εκπαίδευσης ενός μοντέλου. Αυτά περιγράφονται χρησιμοποιώντας στοιχεία της παρακάτω εικόνας για να γίνουν πλήρως κατανοητά:

- **Higher start:** η αρχική απόδοση του target task, χρησιμοποιώντας μόνο τη μεταφερόμενη γνώση από το source task, πριν ακόμη εκπαιδευτεί περαιτέρω, σε σύγκριση με την αρχική απόδοση ενός τυχαία αρχικοποιημένου μοντέλου είναι υψηλότερη. Έτσι, αντί να ξεκινήσει από την αρχή η διαδικασία εκπαίδευσης ή εξαγωγής χαρακτηριστικών, για παράδειγμα, διατηρούνται πρότυπα και μοτίβα που αποκτήθηκαν από την επίλυση της «πρώτης» εργασίας. Αυτό πραγματοποιείται μέσω της μεταφοράς των τιμών των βαρών, όπως έχουν προκύψει μετά από εκπαίδευση στο αρχικό σύνολο δεδομένων.
- **Higher slope:** ο απαιτούμενος χρόνος εκπαίδευσης του target task είναι πολύ μικρότερος και κατ'επέκταση ο ρυθμός βελτίωσης των «δεξιότητων» του, λαμβάνοντας υπόψη τη μεταφερόμενη γνώση, είναι πολύ μεγαλύτερος σε σχέση με τα αντίστοιχα μεγέθη στην περίπτωση ενός μοντέλου που εκπαιδεύεται από την αρχή.
- **Higher asymptote:** η τελική απόδοση του target task είναι αισθητά υψηλότερη συγκριτικά με την τελική απόδοση χωρίς μεταφορά μάθησης [119].

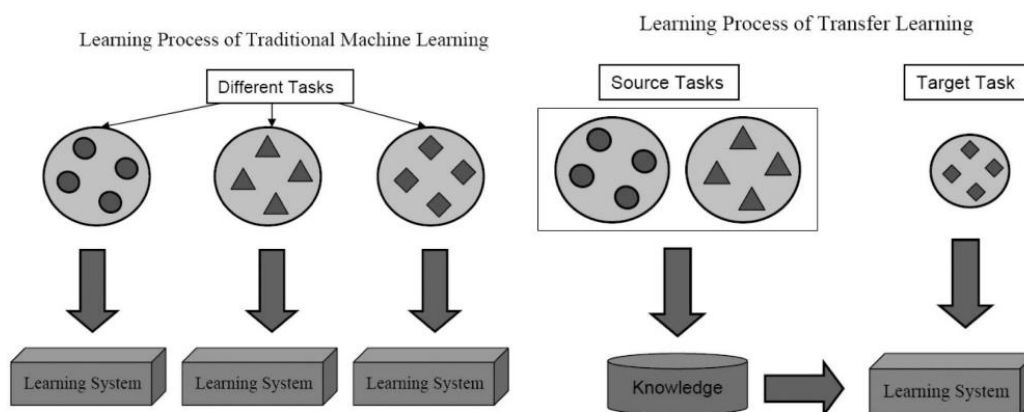


Σχήμα 3.25: Συγκριτική παρουσίαση της απόδοσης ενός δικτύου που χρησιμοποιεί μεταφορά μάθησης, με δίκτυο που δεν χρησιμοποιεί. (Πηγή [119])

Οι εφαρμογές της τεχνικής αυτής σε προβλήματα Μηχανικής Μάθησης είναι ποικίλες. Χαρακτηριστικά παραδείγματα αποτελούν η Ιατρική, οι Συγκοινωνίες, διάφορες Κοινωνικές και Οικονομικές εφαρμογές ή ακόμη και χρήση της για ανάπτυξη Συστημάτων Συστάσεων.

Εν γένει, στόχος των αλγορίθμων ML είναι να αποδίδουν βέλτιστες προβλέψεις σε άγνωστα δεδομένα και αυτό το επιτυγχάνουν μέσω της εκπαίδευσης πάνω σε ένα επισημασμένο σύνολο δεδομένων (labeled data). Επομένως, γίνεται κατανοητό ότι προκειμένου αυτή η εκπαίδευση να είναι αποτελεσματική και να προκύψει ένα μοντέλο

που θα έχει δυνατότητα γενίκευσης βάσει της αποκτηθείσας γνώσης, θα πρέπει να είναι διαθέσιμη πληθώρα τέτοιων δειγμάτων, δεδομένης της πολυπλοκότητας αυτών των μοντέλων, αλλά και προς αποφυγή προβλημάτων υπερπροσαρμογής (overfitting). Κάτι τέτοιο, δυστυχώς, δεν είναι εύκολο να συμβεί σε real-world εφαρμογές, καθώς η συλλογή τους μπορεί να αποδειχθεί εξαιρετικά χρονοβόρα, αλλά και κοστοβόρα. Σε ορισμένες περιπτώσεις, μάλιστα, ιδιαίτερος στη σύγχρονη εποχή, όπου οι νόμοι περί ιδιωτικότητας και προστασίας προσωπικών δεδομένων είναι πολύ πιο αυστηροί, η συλλογή τέτοιων δειγμάτων καθίσταται ακόμη δυσκολότερη, έως και αδύνατη [118]. Συνεπώς, είναι προφανές ότι σε σύγκριση με την παραδοσιακή διαδικασία εκπαίδευσης ενός μοντέλου μηχανικής μάθησης, όπου μπορεί να παρουσιαστούν τα παραπάνω προβλήματα, η χρήση της Μεταφοράς Μάθησης φαίνεται να υπερτερεί σημαντικά.



Σχήμα 3.26: Παραδοσιακή Μηχανική Μάθηση (αριστερά) και Μεταφορά Μάθησης (δεξιά). (Πηγή [118])

Συνοψίζοντας, μέσω της μεταφοράς μάθησης, επωφελούμαστε από τη γνώση που έχει ήδη αποκτηθεί από άλλες συναφείς εργασίες και τη χρησιμοποιούμε για την επίλυση νέων προβλημάτων. Αυτή η προσέγγιση μας επιτρέπει να αποφύγουμε την ανάγκη για μεγάλο όγκο ετικετοποιημένων δεδομένων και να επιτύχουμε ταχύτερα και πιο αποδοτικά την εκπαίδευση των εκάστοτε μοντέλων μηχανικής μάθησης. Τέλος, με τη βοήθεια της τεχνικής αυτής, μπορούμε να αντιμετωπίσουμε προβλήματα όπως η έλλειψη διαθέσιμων δεδομένων για εκπαίδευση, η ανάπτυξη εκ νέου πολύπλοκων μοντέλων, καθώς και οι αλλαγές στη κατανομή των δεδομένων με το πέρασμα του χρόνου.



# Μέρος II

Πρακτικό Μέρος

---



## Κεφάλαιο 4

### 4 Ανάλυση και Σχεδίαση

---

Στο κεφάλαιο αυτό γίνεται αναφορά στο σύνολο δεδομένων που χρησιμοποιήθηκαν για την υλοποίηση της εργασίας, καθώς και στη μεθοδολογία που ακολουθήθηκε. Αρχικά, δίνονται πληροφορίες για τα δεδομένα που επιλέχθηκαν και την επεξεργασία που υπέστησαν. Στη συνέχεια, περιγράφονται τα υπολογιστικά και προγραμματιστικά εργαλεία που βοήθησαν στην εκπόνηση του πρακτικού μέρους της διπλωματικής και, τέλος, παρουσιάζονται οι διαφορετικές προσεγγίσεις σχετικά με τις αρχιτεκτονικές των νευρωνικών δικτύων. Πραγματοποιήθηκαν διάφοροι πειραματισμοί, τόσο με απλά Συνελκτικά Νευρωνικά Δίκτυα που δημιουργήθηκαν *from scratch*, όσο και με προ-εκπαιδευμένα μοντέλα.

#### 4.1 Δεδομένα: Ανάλυση και Προεπεξεργασία

##### 4.1.1 Σύνολο Δεδομένων

Το σύνολο δεδομένων (dataset) που χρησιμοποιήθηκε ως βάση προέρχεται από το Ερευνητικό πρόγραμμα «Θράκη – Μακεδονία» του Συλλόγου «Οι Φίλοι της Μουσικής» που έλαβε χώρα κατά το χρονικό διάστημα 1995 – 2000 [7]. Η ερευνητική ομάδα που συστάθηκε για το σκοπό αυτό, η οποία αποτελούνταν από ιστορικούς, ανθρωπολόγους, εθνομουσικολόγους, θεωρητικούς του χορού, άλλους ειδικούς συνεργάτες, καθώς και νέους ερευνητές, όρισε ως στόχο της να μελετήσει το τραγούδι, τη μουσική και το χορό σε ιστορική και πολιτισμική προοπτική. Ηχογραφήθηκαν περισσότερα από 2.500 τραγούδια, οργανικοί σκοποί και δείγματα λόγου σε 116 ψηφιακές ταινίες διάρκειας 130 ωρών και 1.000 κασέτες αναλογικού ήχου με συνεντεύξεις. Επίσης, συγκεντρώθηκε οπτικό υλικό σε βίντεο διάρκειας περίπου 400 ωρών και 12.000 φωτογραφίες και διαφάνειες. Για την εν λόγω εργασία χρησιμοποιήθηκαν μόνο τα κινησιογράμματα, τα οποία καταγράφηκαν για ορισμένους από τους χορούς που μελετήθηκαν στο ερευνητικό αυτό πρόγραμμα.

Πρόκειται στο σύνολο για 133 εικόνες μορφής JPEG, καθεμία από τις οποίες περιλαμβάνει ένα ή περισσότερα κινησιογράμματα, τα οποία αναπαριστούν χορευτικά μοτίβα παραδοσιακών χορών της Θράκης και της Μακεδονίας από 65 διαφορετικές πόλεις και ακολουθούν τη μέθοδο καταγραφής του Laban (Labanotation). Το κάθε αρχείο έχει υποστεί επεξεργασία στα πλαίσια της παρούσας εργασίας. Πιο συγκεκριμένα, από τις αρχικές εικόνες έχουν δημιουργηθεί νέες, καθεμία από τις οποίες περιλαμβάνει ένα μόνο κινησιογράμμα που περιγράφει τον εκάστοτε χορό. Επειδή, όμως, στις περισσότερες εικόνες περιέχονταν περισσότερα του ενός κινησιογράμματα, στην πλειονότητα των χορών υπάρχουν τουλάχιστον 2 με 3 παραλλαγές τους. Μάλιστα, στους πιο «κλασικούς» χορούς, όπως ο Ζωναράδικος, ο Συγκαθιστός κ.α. δημιουργήθηκαν αρκετές παραλλαγές τους, 33 και 16 αντιστοίχως. Στα νέα αρχεία έχουν διατηρηθεί μόνο όσα ήταν σχετικά με το κινησιογράμμα αυτό καθ' αυτό και σε πολλές περιπτώσεις τα διαφορετικά κινησιογράμματα ή τα αποσπάσματα αυτών, που υπήρχαν σε μία εικόνα, έχουν συνδυαστεί σε ένα ενιαίο, με σκοπό να αναπαραστήσουν ολοκληρωμένες παρτιτούρες χορού, μη διακεκομμένες. Όλες οι τροποποιήσεις έχουν γίνει με στόχο να προκύψουν δεδομένα ομοιόμορφης δομής, ώστε να είναι όσο το δυνατόν ευκολότερο να χρησιμοποιηθούν στο Πρακτικό Μέρος της εργασίας και να παραχθούν έγκυρα αποτελέσματα.

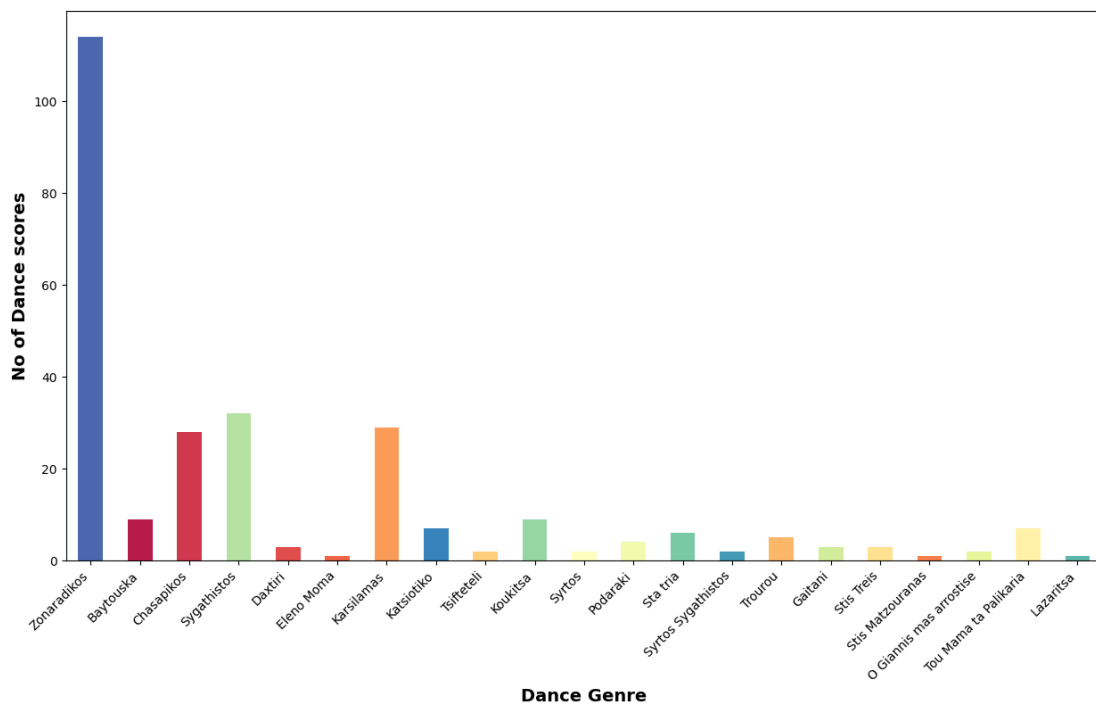
Τα τελικά δεδομένα, μετά την επεξεργασία, αποτελούνται από 271 πλαίσια διαφορετικού μεγέθους εικονοστοιχείων το καθένα, τα οποία δεν είναι επισημασμένα (labeled). Το μέγεθος του συνόλου είναι 20,60 MB και η μορφή είναι JPEG. Υπάρχουν 21 διαφορετικά γένη χορών: Ζωναράδικος, Συρτός, Καρσιλαμάς, Συγκαθιστός, Συρτός Συγκαθιστός, Παϊντούσκα ή Μπαϊντούσκα, Χασάπικος, Κατσιώτικο, Ντάχτιρι, Ποδαράκι, Ελένο Μόμα, Κουκίτσα, Στα τρία, Τροϋρου, Τσιφτετέλι, Γαϊτάνι - Κουσευτός, Στις Τρεις ή Τρεις, Στης Ματζουράνας, Ο Γιάννης μας αρρώστησε, Του Μαμά Τα Παλληκάρια, Λαζαρίτσα. Μάλιστα, κάποιες από αυτές τις κατηγορίες παραδοσιακών χορών της Θράκης χωρίζονται σε επιμέρους είδη. Στο σύνολο δεδομένων που χρησιμοποιούμε υπάρχουν τα εξής επιμέρους είδη χορών:

- **Ζωναράδικος:** Του κνα, Ντοντόρου, Σιγανό, Κουλουριαστό, Ξεσυρτός, Λεν' Ηρθι Μάης, Ταπεινός, Ν' αραδιαστείτε στου χορού, Σουλτάνα Σουφλιουτούδα, Σούστα, Γκίκνα, Συμπεθεριάτικος, Ζαβρατένο, Ζωναράδικος Τσέστος, Για βγείτε ν' αγναντέψετε, Κλωστρός, Κουρτέσα, Ντρίστα, Νταϊσάλ Ρους, Τσατσάκι, Ζερβός, Γονατιστός.
- **Χασάπικος:** Χασαπιά, Αγαπό Μάνα, Κλάψετέ με φανταράκια, οι φιληνάδες, Μίλα Χρυσάνθη, Εχθές το βράδυ καλέ μάνα, Διαμάντω.

- **Συγκαθιστός:** Νταουλτζίδικος, Κετσέκι, Κιρκ' λι π' λάκια, Συγκάθισμα, Συγκα' στό του γάμου.
- **Καρσιλαμάς:** Συγκάθισμα, Το Λενάκι πλάθει πίτα, Στο χωριό μας το καμάρι έχει ένα παλικάρι.
- **Τσιφτετέλι:** Κονιάλης, Κουτσός.
- **Συρτός:** Καλαματιανός, Σταματία.
- **Λαζαρίτσα:** Το βράδυ Άννα, Βαγγελίτσα, Τσομπανάκος.

Καθεμία από τις παραπάνω 21 κατηγορίες περιλαμβάνει διαφορετικό αριθμό κινήσιογραμμάτων. Ο Ζωναράδικος περιέχει 114 κινήσιογράμματα (πάνω από το 40% των συνολικών δειγμάτων), ο Συγκαθιστός 32, ο Καρσιλαμάς 29, ο Χασάπικος 28, η Μπαϊντούσκα 9, ο Συρτός 9, το Τσιφτετέλι 7, ο Γιάννης μας αρρώστησε 7, ο Συρτός Συγκαθιστός 6, το Κατσιώτικο 5, ο Στα τρία 4, το Ντάχτρι 3, η Λαζαρίτσα 3, ο Στις τρεις 3, το Ποδαράκι 2, ο Στις Ματζουράνας 2, η Τροῦρου 2, η Ελένο Μόμα 1, το Γαϊτάι 1 και, τέλος, ο Του Μαμά τα παλληκάρια 1.

**Visualization of the number of dance scores within each dance genre**



*Σχήμα 4.1: Πλήθος κινήσιογραμμάτων σε κάθε γένος χορού.*

Παραπάνω φαίνεται το διάγραμμα, όπου οπτικοποιήσαμε τα όσα αναφέρθηκαν, ώστε να γίνουν πιο εύκολα αντιληπτά. Είναι οφθαλμοφανές, λοιπόν, ότι πρόκειται για ένα

ανισόρροπο σύνολο δεδομένων (*imbalanced dataset*), όπου η κλάση πλειοψηφίας (*majority class*) και, μάλιστα, με διαφορά είναι η κατηγορία του *Ζωναράδικου*. Όπως καταλαβαίνουμε, αυτό θα προσθέσει μια επιπλέον δυσκολία στο πρόβλημα της συσταδοποίησης, δημιουργώντας προβλήματα, όπως είναι το *over-representation*, με την «κυρίαρχη» κλάση να υπερτερεί σε ένα ή και περισσότερα clusters, ενώ όσον αφορά τις κλάσεις μειοψηφίας (*minority classes*) να αγνοούνται και πολλές φορές να συγχωνεύονται σε μεγαλύτερες συστάδες, οπότε να καθίσταται δύσκολος ο εντοπισμός των μοναδικών χαρακτηριστικών τους. Επιπροσθέτως, αυτή η ανισορροπία μπορεί να προκαλέσει προβλήματα και στην αξιολόγηση της ποιότητας της συσταδοποίησης, με τις μετρικές που χρησιμοποιούνται να επηρεάζονται συχνά λόγω της άνισης κατανομής των δεδομένων σε κλάσεις.

Παρότι το σύνολο δεδομένων είναι μικρού μεγέθους παρουσιάζει ορισμένα βασικά πλεονεκτήματα. Καταρχήν, τα κινησιογράμματα που περιλαμβάνει είναι γραμμένα με «ομοιόμορφο» τρόπο, δηλαδή χωρίς να παρεκκλίνει ιδιαίτερος το ένα από το άλλο, όσον αφορά το στυλ της σημειογραφίας, κάτι το οποίο είναι θετικό για την αξιολόγηση των αποτελεσμάτων. Να επισημανθεί ότι τα συστήματα καταγραφής χορού, όπως το Labanotation σε αυτήν την περίπτωση, αποτελούν μεν μια «κοινή γλώσσα» για την αποτύπωση της χορευτικής κίνησης, όμως ο/η κάθε σημειογράφος με βάση τις συνθήκες, όπως για παράδειγμα είναι το κοινό στο οποίο απευθύνεται ή η πολυπλοκότητα του χορευτικού συνδυασμού [4], επιλέγει τον τρόπο με τον οποίο θα δημιουργήσει τις παρτιτούρες του χορού. Έτσι, γίνεται αντιληπτό ότι δεν πρόκειται για μια γλώσσα με «αυστηρούς» κανόνες, αλλά για ένα σύστημα πιο περιγραφικό που επιτρέπει στον δημιουργό να χρησιμοποιήσει τα στοιχεία που θεωρεί εκείνος απαραίτητα αναλόγως την περίπτωση. Οπότε, το γεγονός ότι τα συγκεκριμένα κινησιογράμματα έχουν γραφεί ακολουθώντας ένα κοινό πρότυπο είναι ιδιαίτερος σημαντικό για την έρευνά μας. Επιπλέον, κάθε χορός συνοδεύεται από τα μεταδεδομένα του, μια καρτέλα (σε μορφή .pdf αρχείου), η οποία παρέχει συμπληρωματικές πληροφορίες σχετικά με το είδος του χορού, τον τόπο από όπου προέρχεται, τη μορφή της κινητικής ενότητας, το σχήμα του, καθώς και το μουσικό είδος, το μέτρο, τη ρυθμική αγωγή κ.ο.κ. Τα στοιχεία αυτά μας βοηθούν στην εις βάθος μελέτη των κινησιογραμμάτων και στην ποιοτική αξιολόγηση των αποτελεσμάτων που προκύπτουν από τη συσταδοποίηση. Τέλος, να ληφθεί υπόψη ότι πρόκειται για ένα dataset δημόσιο, το οποίο αποτελείται από αρχεία πολύτιμα, μεγάλης πολιτισμικής αξίας που εν γένει δεν είναι εύκολη η εύρεσή τους, καθώς τέτοιου είδους κινησιογράμματα στην πλειοψηφία τους δεν είναι διαθέσιμα για το ευρύ κοινό.

Ακόμη, στα πλαίσια της διπλωματικής εργασίας, δημιουργήθηκε ένα .csv αρχείο, όπου καταχωρήθηκαν τα 271 κινησιογράμματα – εικόνες αντιστοιχιζόμενα το καθένα με 1 από τα 21 διαφορετικά γένη χορού. Αυτό πραγματοποιήθηκε, με τη βοήθεια των μεταδεδομένων που συνόδευαν κάθε αρχική εικόνα προερχόμενη από το ερευνητικό αυτό πρόγραμμα. Το τελικό αρχείο που προέκυψε αποδεικνύεται εξαιρετικά χρήσιμο για τη συνέχεια, όσον αφορά την αξιολόγηση των αποτελεσμάτων.

#### **4.1.2 Προεπεξεργασία Δεδομένων**

Η προεπεξεργασία των δεδομένων και πιο συγκεκριμένα, στην περίπτωση μας των εικόνων, αποτελεί μια εξαιρετικά χρήσιμη και, μάλλον, απαραίτητη διαδικασία στον τομέα της Όρασης Υπολογιστών και της Αναγνώρισης Προτύπων. Περιλαμβάνει αρκετές τεχνικές που χρησιμοποιούνται για να ενισχύσουν την οπτική ποιότητα των εικόνων, κάνοντάς τις πιο ευκρινείς, καθαρότερες. Με αυτόν τον τρόπο, γίνονται περισσότερο εμφανή και τα χαρακτηριστικά τους, δίνοντας περισσότερη πληροφορία. Απώτερος σκοπός είναι τα βελτιωμένα αποτελέσματα των αλγορίθμων μηχανικής μάθησης, όπου χρησιμοποιούνται σαν είσοδος.

Στην παρούσα εργασία, έγιναν δοκιμές με αρκετές μεθόδους προεπεξεργασίας που σχετίζονται με τη βελτίωση των δεδομένων - εικόνων, όπως μείωση του θορύβου (Noise Reduction), τόνωση (Sharpening), ενίσχυση χρώματος & αντίθεσης (Color & Contrast Enhancement), παρόλα αυτά τα αποτελέσματα δεν ήταν τα αναμενόμενα, όχι μόνο δε βελτιώθηκαν, αλλά οι επιδόσεις των μοντέλων που χρησιμοποιήθηκαν στο πειραματικό κομμάτι της εργασίας ήταν μειωμένες. Έτσι, αποδεικνύεται για ακόμη μια φορά ότι η χρησιμότητα αυτών των εργαλείων έγκειται στη φύση των δεδομένων και, φυσικά, τους στόχους του προβλήματος. Τελικά, από το πεδίο αυτό χρησιμοποιήθηκαν τρεις τεχνικές: Αλλαγή μεγέθους εικόνων (Resizing & Rescaling), Εξαγωγή χαρακτηριστικών (Feature Extraction) και Μείωση Διαστατικότητας (Dimensionality Reduction).

##### **4.1.2.1 Αλλαγή μεγέθους εικόνων**

Για την επεξεργασία των εικόνων χρησιμοποιούνται εργαλεία από τη βιβλιοθήκη Keras [120], στην οποία θα γίνει εκτενέστερη αναφορά παρακάτω. Πρώτο βήμα αποτελεί η φόρτωση των εικόνων σε συγκεκριμένες διαστάσεις (target\_size). Αυτό είναι απαραίτητο, καθώς το κάθε CNN δέχεται ως είσοδο εικόνες με σταθερό και προκαθορισμένο μέγεθος, οπότε αναλόγως με τις απαιτήσεις του κάθε μοντέλου, ορίζονται

και οι διαστάσεις. Οι εικόνες, αρχικά, φορτώνονται ως αντικείμενα PIL, λόγω της χρησιμοποιούμενης μεθόδου. Ωστόσο, αυτή η μορφή δε θα μας φανεί χρήσιμη, οπότε τις μετατρέπουμε σε εύκολα επεξεργάσιμους πίνακες (numpy array). Οι πίνακες που προκύπτουν είναι τρισδιάστατοι με δομή: (Μήκος, Πλάτος, Αριθμός Καναλιών). Όμως τα νευρωνικά δίκτυα λειτουργούν με *batches*, δηλαδή υποσύνολα, «δεμάτια» δεδομένων, οπότε οι πίνακες αυτοί θα πρέπει να περιλάβουν μία ακόμη διάσταση που θα ενημερώνει το μοντέλο πόσα *batches*, με άλλα λόγια πόσες εικόνες θα δεχτούν κάθε φορά. Έτσι, αλλάζουμε για τελευταία φορά το μέγεθος των εικόνων, μετατρέποντάς τες σε πίνακες με την εξής δομή: (Αριθμός δειγμάτων, Μήκος, Πλάτος, Κανάλια).

#### 4.1.2.2 Εξαγωγή χαρακτηριστικών

Ακολουθεί το σημαντικότερο κομμάτι της προεπεξεργασίας του συνόλου δεδομένων, που δεν είναι άλλο παρά η εξαγωγή χρήσιμων χαρακτηριστικών από τα κινησιογράμματα. Αυτό επιτυγχάνεται με τη βοήθεια των Συνελκτικών Νευρωνικών Δικτύων, όπως έχει αναλυθεί σε προηγούμενο κεφάλαιο, είτε αυτά είναι προ-εκπαιδευμένα, είτε όχι, αναμένοντας, φυσικά, τις αντίστοιχες επιδόσεις. Σε κάθε περίπτωση, για να λάβουμε και να μπορούμε να χρησιμοποιήσουμε εν συνεχεία τα σημαντικά εκείνα χαρακτηριστικά από κάθε εικόνα, θα πρέπει να τροποποιήσουμε ελαφρώς τη δομή των δικτύων. Πιο συγκεκριμένα, λόγω του ότι τα CNN συχνά χρησιμοποιούνται σε εφαρμογές ταξινόμησης εικόνων, το επίπεδο εξόδου τους αποτελείται από πλήρως συνδεδεμένα layers, τα οποία αναδεικνύουν την κλάση της εκάστοτε εισόδου. Στην παρούσα εργασία, ωστόσο, ασχολούμαστε με ένα πρόβλημα συσταδοποίησης και όχι ταξινόμησης, οπότε η έξοδος αυτή των νευρωνικών δικτύων δε θα μπορούσε κάπως να αξιοποιηθεί. Για αυτό το λόγο, εφόσον σκοπός μας είναι να εξάγουμε απλώς τα σημεία ενδιαφέροντος από το σύνολο δεδομένων, αφαιρούμε το τελευταίο *fully-connected layer*, που ευθύνεται για την κατηγοριοποίηση και έτσι, λαμβάνουμε τα χαρακτηριστικά από το αμέσως προηγούμενο στρώμα του δικτύου μας. Πρόκειται για μια συνήθη τεχνική που χρησιμοποιείται σε αυτές τις περιπτώσεις.

Ως εξαγωγείς χαρακτηριστικών στην εν λόγω εργασία χρησιμοποιήθηκαν τόσο δίκτυα που δημιουργήθηκαν *from scratch* και, επομένως, δεν είχαν εκπαιδευτεί πρωτίτερα σε κάποιο άλλο σύνολο δεδομένων, όσο και δίκτυα προ-εκπαιδευμένα σε τεράστια σύνολα δεδομένων. Έτσι, αξιοποιείται και η μέθοδος του transfer learning, μέσω των συχνά χρησιμοποιούμενων μοντέλων σε προβλήματα ανάλυσης εικόνων: VGG16, VGG19, ResNet50, InceptionV3 και Xception, θέλοντας να μελετήσουμε διαφορετικά είδη CNN,



να εξετάσουμε τον τρόπο λειτουργίας τους, αλλά και τις επιδόσεις τους στο συγκεκριμένο πρόβλημα συσταδοποίησης. Για την εξαγωγή χαρακτηριστικών χρησιμοποιήθηκε η μέθοδος *predict*, δίνοντάς μας το αντίστοιχο διάνυσμα χαρακτηριστικών για κάθε εικόνα – παρτιτούρα. Τα διανύσματα αυτά αποθηκεύτηκαν σε ένα λεξικό, όπου το όνομα του κάθε χορού (και πιο συγκεκριμένα το όνομα του κάθε αρχείου) αποτελεί το κλειδί και το διάνυσμα χαρακτηριστικών την αντίστοιχη τιμή.

#### 4.1.2.3 Ανάλυση Κυρίων Συνιστωσών

Στο επόμενο βήμα, το τελευταίο πριν τη συσταδοποίηση των κινησιογραμμάτων, επιλέχθηκε να γίνει μείωση της διαστατικότητας των δεδομένων, μέσω της πολύ γνωστής τεχνικής PCA, η οποία έχει, επίσης, αναλυθεί στο θεωρητικό μέρος της εργασίας. Έτσι, τα *feature maps* δε χρησιμοποιούνται κατευθείαν ως είσοδος στον εκάστοτε αλγόριθμο συσταδοποίησης, αλλά υφίστανται σε πρώτη φάση μια μείωση των διαστάσεών τους. Αυτό έχει σαν στόχο την ελάττωση του όγκου της πληροφορίας και κατ' επέκταση τη χρονική, υπολογιστική αλλά και ποιοτική βελτιστοποίηση των τελικών αποτελεσμάτων.

Είναι προφανές ότι αυτή η διαδικασία μείωσης της διαστατικότητας δε θα πρέπει να βασίζεται στην τυχαιότητα, δηλαδή δε θα ήταν καθόλου ωφέλιμο, απλώς να μικραίναμε το μέγεθος των παραγόμενων διανυσμάτων, καθώς μπορεί να χάνονταν σημαντικά χαρακτηριστικά, απαραίτητα για την υλοποίηση της συσταδοποίησης που ακολουθεί. Για αυτόν, ακριβώς, το λόγο, επιλέχθηκε η μέθοδος Ανάλυσης Κυρίων Συνιστωσών, κατά την οποία, με βάση το αρχικό σύνολο δεδομένων, δημιουργείται ένα νέο, μικρότερο του πρώτου, το οποίο, όμως, διατηρεί όσο το δυνατόν περισσότερη πληροφορία. Η απόφαση, όσον αφορά τον αριθμό των κύριων συνιστωσών (*principal components*) δε λαμβάνεται αυτοματοποιημένα, αλλά ορίστηκε από εμάς, λαμβάνοντας υπόψιν τα τελικά αποτελέσματα. Έγιναν πολλές δοκιμές, ώστε να βρεθεί ο κατάλληλος αριθμός συνιστωσών για το κάθε μοντέλο. Κάθε φορά, επιλεγόταν ο μικρότερος δυνατός, με την προϋπόθεση να διατηρούνται τα σημαντικά σημεία ενδιαφέροντος των δεδομένων.

## 4.2 Σχεδιασμός και Περιγραφή υλοποίησης

Μετά την ολοκλήρωση της φάσης επεξεργασίας του συνόλου δεδομένων, ακολουθεί η συσταδοποίηση, βασική θεματική της παρούσας διπλωματικής. Βασιζόμενοι στα χαρακτηριστικά που είχαν εξαχθεί από τα προηγούμενα βήματα, προσεγγίσαμε το πρόβλημα της συσταδοποίησης, χρησιμοποιώντας έναν από τους πιο συχνά

χρησιμοποιούμενους αλγόριθμους, τον αλγόριθμο των k-μέσων, γνωστός ως k-means. Όπως έχει αναλυθεί στο τρίτο κεφάλαιο της εργασίας, στόχος του αλγορίθμου αυτού είναι να ομαδοποιήσει τα δεδομένα σε k συστάδες με βάση την ομοιότητά τους. Ο αριθμός των συστάδων, ωστόσο, δεν είναι προκαθορισμένος, ούτε επιλέγεται αυτόματα από τον ίδιο τον αλγόριθμο με βάση τα δεδομένα, αλλά ορίζεται από τον χρήστη. Για αυτό το λόγο, χρήσιμη αποδεικνύεται πολλές φορές η χρήση τεχνικών προσδιορισμού του k. Ωστόσο, στην περίπτωση μας, λόγω του ότι είχαμε ήδη δημιουργήσει το προαναφερθέν .csv αρχείο προς διευκόλυνσή μας για την αξιολόγηση των αποτελεσμάτων, υπήρχε πρότερη γνώση σχετικά με τον ακριβή αριθμό συστάδων, στις οποίες ιδανικά θα έπρεπε να ομαδοποιηθούν τα δεδομένα. Οπότε, ορίστηκε  $k = 21$ , εφόσον είχαμε 21 διαφορετικά γένη χορού. Ωστόσο, αξίζει να σημειωθεί ότι έγιναν πειραματισμοί και με διαφορετικούς αριθμούς συστάδων, κάτι το οποίο θα συζητηθεί στο επόμενο κεφάλαιο.

Σε αυτό το σημείο, σημαντικό κρίνεται να αναφερθεί ότι ο αλγόριθμος k-means, προκειμένου να ομαδοποιήσει τα δεδομένα, υπολογίζει κάθε φορά την απόσταση δείγματος - centroid και το κατατάσσει στην ομάδα με το πλησιέστερο προς το δείγμα κέντρο. Στόχος, δηλαδή, είναι να ελαχιστοποιείται αυτή η απόσταση, που μετρείται ως inertia ή αλλιώς within-cluster sum-of-squares. Στην πραγματικότητα, η μετρική αυτή ταυτίζεται με την Ευκλείδεια απόσταση (Euclidean distance) υψωμένη στο τετράγωνο και δίνεται από τον τύπο:

$$Inertia(C) = \sum_{x \in C} \|x - \mu(C)\|^2$$

όπου το  $x$  αναπαριστά το κάθε δείγμα, το  $C$  αναπαριστά την κλάση και το  $\mu(C)$  το κέντρο - centroid της εκάστοτε κλάσης.

Πειραματισμοί έγιναν και με άλλες μετρικές, όπως manhattan, chebyshev, canberra, καθώς και chi-square distance, ωστόσο τα αποτελέσματα έδειξαν ότι η μετρική που χρησιμοποιεί by default ο αλγόριθμος k-means αποτελεί τη βέλτιστη επιλογή για το συγκεκριμένο σύνολο δεδομένων.

#### 4.2.1 Εργαλεία

Η γλώσσα προγραμματισμού που χρησιμοποιήθηκε στην παρούσα εργασία είναι η Python3. Πρόκειται για μια υψηλού επιπέδου γλώσσα, η οποία χρησιμοποιείται από πολύ μεγάλο αριθμό προγραμματιστών, χάρη στη συντακτική της απλότητα, τη διαθεσιμότητά

της στο ευρύ κοινό, καθώς και το μεγάλο εύρος βιβλιοθηκών που διαθέτει. Είναι η γλώσσα που συναντάται, κατά κόρον, σε τέτοιου είδους προβλήματα, Βαθιάς Μάθησης.

Όσον αφορά την υλοποίηση, καθώς και αξιολόγηση των μοντέλων, πραγματοποιήθηκαν με τη βοήθεια των βιβλιοθηκών Tensorflow [121], Keras [120], το οποίο λειτουργεί ως διεπαφή για τη βιβλιοθήκη Tensorflow, όπως επίσης και Scikit-Learn [122]. Πρόκειται για βιβλιοθήκες ανοικτού κώδικα, ευρέως χρησιμοποιούμενες στο πεδίο της Μηχανικής Μάθησης, κατάλληλες για τη χρήση νευρωνικών δικτύων. Άλλες αριθμητικές και επιστημονικές βιβλιοθήκες της python που χρησιμοποιήθηκαν στα πλαίσια της εργασίας είναι η NumPy [123], η pandas [124], η cv2 (OpenCV) [125], καθώς και η βιβλιοθήκη σχεδίασης και οπτικοποίησης Matplotlib [126].

Τέλος, να αναφερθεί ότι το πειραματικό μέρος της εργασίας, η υλοποίηση του κώδικα και η εξαγωγή των αποτελεσμάτων πραγματοποιήθηκαν με τη βοήθεια του Google Colaboratory. Πρόκειται για ένα ευρέως χρησιμοποιούμενο υπολογιστικό περιβάλλον που έχει το πλεονέκτημα ότι διατίθεται δωρεάν, φυσικά με ορισμένους περιορισμούς, όμως αποτελεί μια πολύ καλή επιλογή, για χρήστες της γλώσσας Python, παρέχοντας πρόσβαση σε GPUs.

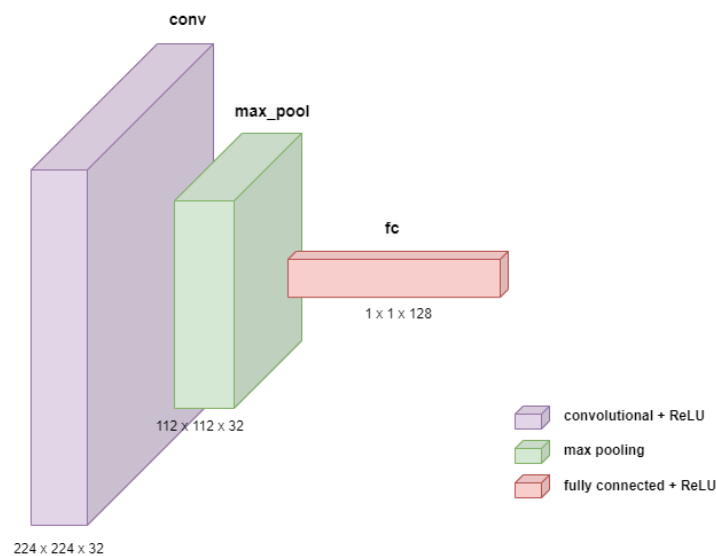
## 4.2.2 Αρχιτεκτονικές Νευρωνικών Δικτύων

Στην υποενότητα αυτή γίνεται αναφορά σε αρχιτεκτονικές συνελκτικών δικτύων που χρησιμοποιήθηκαν στα πλαίσια της παρούσας εργασίας, με σκοπό να γίνει κατανοητή η δομή και τα ιδιαίτερα χαρακτηριστικά τους. Πιο συγκεκριμένα, θα περιγραφούν, αρχικά, απλά μοντέλα που δημιουργήθηκαν από το μηδέν (from scratch) για τις ανάγκες του συγκεκριμένου προβλήματος, και, στη συνέχεια, σύγχρονες προσεγγίσεις δικτύων, οι οποίες αξιοποιήθηκαν με τη βοήθεια της μεταφοράς μάθησης. Οι αρχιτεκτονικές που παρουσιάζονται αποτελούν συχνά χρησιμοποιούμενα μοντέλα, που επικράτησαν λόγω της αποτελεσματικότητάς τους, δηλαδή της επίτευξης υψηλών επιδόσεων σε πληθώρα σύνθετων προβλημάτων.

### 4.2.2.1 CNN δίκτυα from scratch

Σε πρώτη φάση, δημιουργήθηκαν τρία Συνελκτικά Νευρωνικά Δίκτυα, με απλή δομή, ώστε να μελετηθεί η συμπεριφορά τους στο συγκεκριμένο πρόβλημα. Καθένα από αυτά, αποτελεί στην ουσία μία «εξέλιξη» του προηγούμενου του, έχοντας σαν στόχο να παρατηρήσουμε και την αναμενόμενη πρόοδο στα αποτελέσματα που παρουσιάζονται.

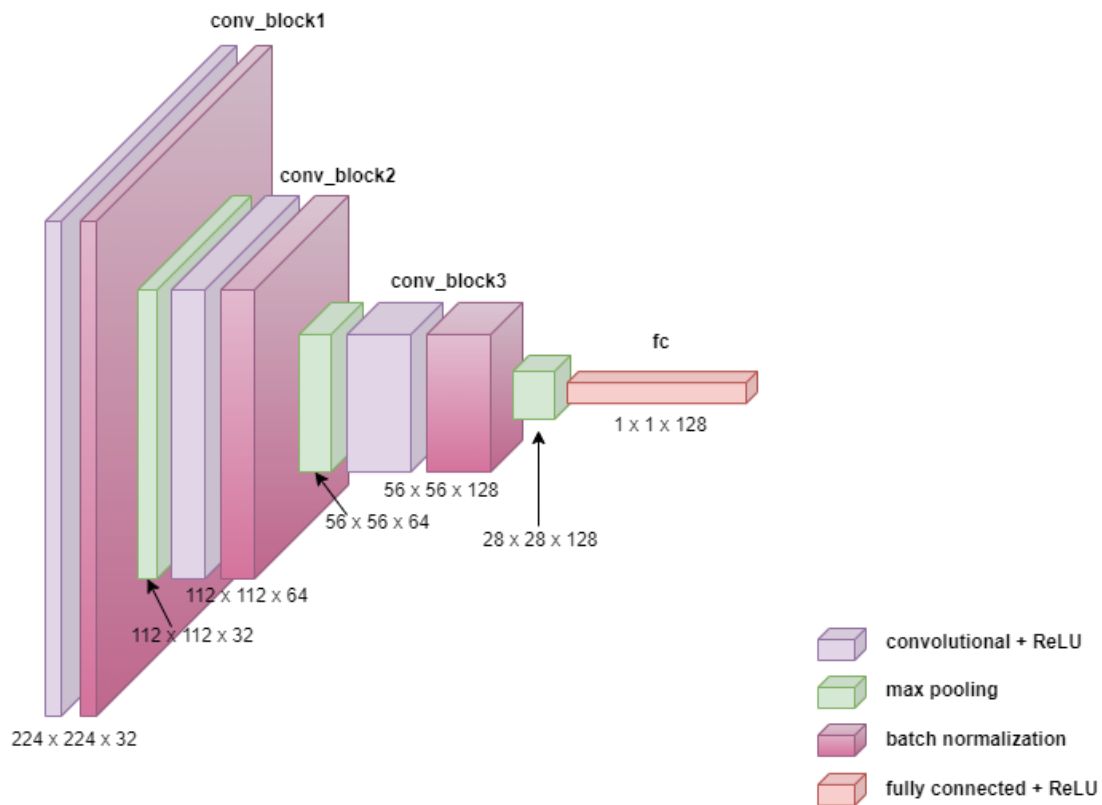
Πιο συγκεκριμένα, το πρώτο μοντέλο, του οποίου η αρχιτεκτονική απεικονίζεται παρακάτω αποτελεί ένα εξαιρετικά απλό δίκτυο που χρησιμοποιείται απλώς για πειραματισμό. Έναυσμα για τη δημιουργία ενός τόσο λιτού μοντέλου ήταν το να εξετάσουμε ποια είναι τα χαμηλότερα σε απόδοση αποτελέσματα που μπορούμε να λάβουμε, ώστε εν συνεχεία να συγκριθούν με τα πολυπλοκότερα δίκτυα. Το μοντέλο αυτό, λοιπόν, αποτελείται από ένα επίπεδο εισόδου που δέχεται RGB εικόνες μεγέθους  $224 \times 224 \times 3$ . Ακολουθεί ένα μόνο συνελκτικό στρώμα που χρησιμοποιεί 32 φίλτρα μεγέθους  $3 \times 3$ , με την παράμετρο padding ενεργή, έτσι ώστε να διατηρείται το μέγεθος των feature maps και να μην έχουμε απώλεια χρήσιμης πληροφορίας. Η συνάρτηση ενεργοποίησης που χρησιμοποιείται στο συνελκτικό επίπεδο είναι η ReLU. Στη συνέχεια, συναντάται ένα max pooling layer, μειώνοντας τις διαστάσεις των χαρτών χαρακτηριστικών στο μισό. Αμέσως μετά, εφαρμόζεται η τεχνική ισοπέδωσης, μέσω της οποίας οι 3D χάρτες μετατρέπονται σε 1D διανύσματα για να ακολουθήσει ένα πλήρως συνδεδεμένο (Dense) layer 128 κόμβων, το οποίο μας δίνει και την έξοδο του δικτύου, η οποία θα χρησιμοποιηθεί στη συνέχεια για τη συσταδοποίηση των εικόνων.



Σχήμα 4.2: Πρώτο CNN μοντέλο from scratch.

Όσον αφορά το δεύτερο μοντέλο, πρόκειται για μια επίσης σχετικά απλή αρχιτεκτονική, όπου αποτελείται από επιμέρους συνελκτικά μπλοκ, χρησιμοποιώντας σαν βάση το μοντέλο που αναλύθηκε παραπάνω. Κάθε συνελκτικό μπλοκ (convolutional block) περιλαμβάνει ένα συνελκτικό επίπεδο με φίλτρα μεγέθους  $3 \times 3$ , ένα στρώμα κανονικοποίησης δέσμης (batch normalization), για τη βελτίωση και τη σταθερότητα του δικτύου, καθώς και ένα επίπεδο υποδειγματοληψίας μεγίστου. Συνολικά αποτελείται από

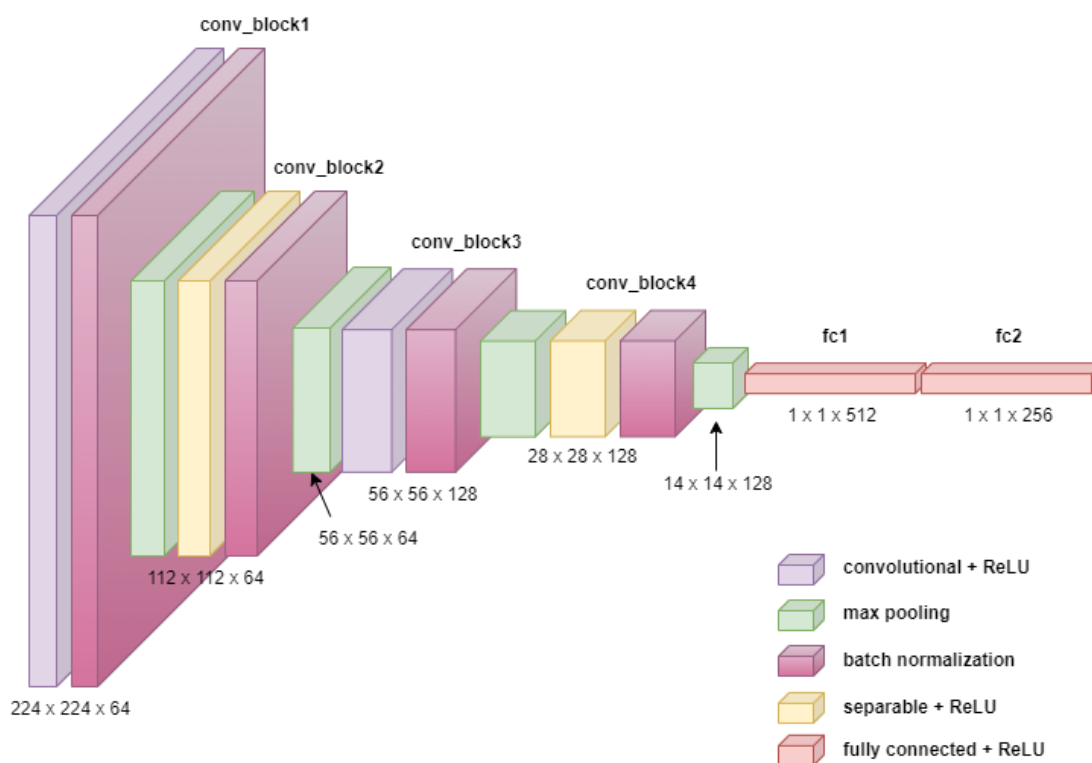
τρία τέτοια μπλοκ, ανάμεσα στα οποία η μόνη διαφορά είναι ότι στο πρώτο έχουμε 32 φίλτρα, στο δεύτερο 64 και στο τρίτο 128. Μετά το flattening, ακολουθεί ένα fully connected layer 128 κόμβων και πάλι, παρέχοντας τα διανύσματα χαρακτηριστικών για την κάθε εικόνα.



Σχήμα 4.3: Δεύτερο CNN μοντέλο from scratch.

Τέλος, στο τρίτο μοντέλο, με το οποίο πειραματιστήκαμε αρκετά, εντάξαμε ένα ακόμη είδος συνελκτικού επιπέδου που υλοποιεί διαχωρίσιμες συνελίξεις και το οποίο ονομάζεται διαχωρίσιμο (separable). Το δίκτυο αυτό επίσης αποτελείται από συνελκτικά μπλοκ, τέσσερα συνολικά αυτή τη φορά. Το πρώτο και το τρίτο μπλοκ περιλαμβάνουν ένα απλό συνελκτικό στρώμα, με φίλτρα μεγέθους  $3 \times 3$ , ένα στρώμα batch normalization, όπως και ένα max pooling layer. Αντιστοίχως, το δεύτερο και τέταρτο μπλοκ περιλαμβάνουν τα ίδια στοιχεία με τη διαφορά ότι το συνελκτικό επίπεδο είναι διαχωρίσιμο. Επίσης, να αναφερθεί ότι τα πρώτα δύο μπλοκ χρησιμοποιούν 64 φίλτρα, ενώ τα δύο τελευταία, 128 φίλτρα. Στη συνέχεια, εφαρμόζεται η μέθοδος flatten και αμέσως, ακολουθούν δύο dense layers, το πρώτο με 512 κόμβους και το δεύτερο με 256. Φυσικά, να αναφέρουμε ότι το μοντέλο αυτό, σε αντίθεση με το πρώτο, δημιουργήθηκε με στόχο να πετύχει βελτιωμένα αποτελέσματα, χωρίς βέβαια, να αυξάνεται κατά πολύ η πολυπλοκότητα. Οπότε, σε αυτήν την περίπτωση έγιναν αρκετές δοκιμές σχετικά με το

πλήθος και το είδος των συνελκτικών επιπέδων, το πλήθος των φίλτρων που θα πρέπει να χρησιμοποιεί καθένα από αυτά, όπως επίσης και με τη διαφορετική διαμόρφωση της αρχιτεκτονικής, βάζοντας δύο συνεχόμενα «διαχωρίσιμα» μπλοκ, όμως τελικά καταλήξαμε σε αυτή τη δομή, η οποία έδινε τα καλύτερα αποτελέσματα.

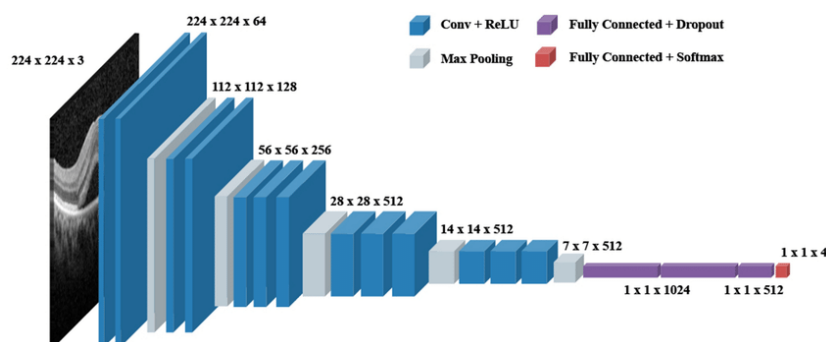


Σχήμα 4.4: Τρίτο CNN μοντέλο from scratch.

#### 4.2.2.2 VGG

Πρόκειται για ένα «κλασικό» Βαθύ Συνελκτικό Νευρωνικό Δίκτυο, το οποίο δημιουργήθηκε από τους Karen Simonyan και Andrew Zisserman, το 2014 [127]. Η ιδιαιτερότητά του έγκειται στο γεγονός ότι βασίζεται στη χρήση των λεγόμενων VGG (Visual Geometry Group) blocks. Πράγματι, οι εμπνευστές του πρότειναν μέσω της αρχιτεκτονικής αυτής, τη δημιουργία ενός δικτύου που περιλάμβανε πολλαπλά μπλοκ, καθένα από τα οποία αποτελούνταν από διαδοχικά συνελκτικά επίπεδα ακολουθούμενα από ένα συγκεντρωτικό επίπεδο. Προτιμήθηκαν πυρήνες μικρών διαστάσεων ( $3 \times 3$ ) για τις συνελίξεις, ενώ σε όλα τα κρυφά επίπεδα χρησιμοποιήθηκε η συνάρτηση ενεργοποίησης ReLU. Στο τέλος του δικτύου συναντώνται 3 πλήρως συνδεδεμένα επίπεδα, εκ των οποίων τα πρώτα δύο έχουν 4096 κανάλια το καθένα, ενώ το τελευταίο 1000, αριθμός ο οποίος αντικατοπτρίζει το πλήθος των κλάσεων.

Υπάρχουν δύο παραλλαγές του μοντέλου αυτού, το VGG16 και το VGG19, τα οποία διαφέρουν ως προς τον αριθμό των επιπέδων που περιλαμβάνουν, με το πρώτο να απαρτίζεται από 16 συνολικά στρώματα, όπως φαίνεται και από το όνομά του, (13 συνελκτικά επίπεδα + 3 πλήρως συνδεδεμένα επίπεδα) και αντίστοιχα το τελευταίο από 19 (16 συνελκτικά επίπεδα + 3 πλήρως συνδεδεμένα επίπεδα). Και τα δύο δέχονται ως είσοδο RGB εικόνες διαστάσεων  $224 \times 224 \times 3$ .



Σχήμα 4.5: Αρχιτεκτονική μοντέλου VGG-16. (Πηγή [128])

Να σημειωθεί ότι αποτελούν δύο από τα πιο διάσημα μοντέλα που κατατέθηκαν στο ILSVRC-2014, πετυχαίνοντας ακρίβεια μεγαλύτερη από 90% στο σύνολο δεδομένων ImageNet. Η αρχιτεκτονική των VGG έχει έκτοτε χρησιμοποιηθεί σαν βάση για πολλά πρωτοποριακά μοντέλα που χρησιμοποιούνται, κατά βάση, για ανίχνευση αντικειμένων και αναγνώριση εικόνων.

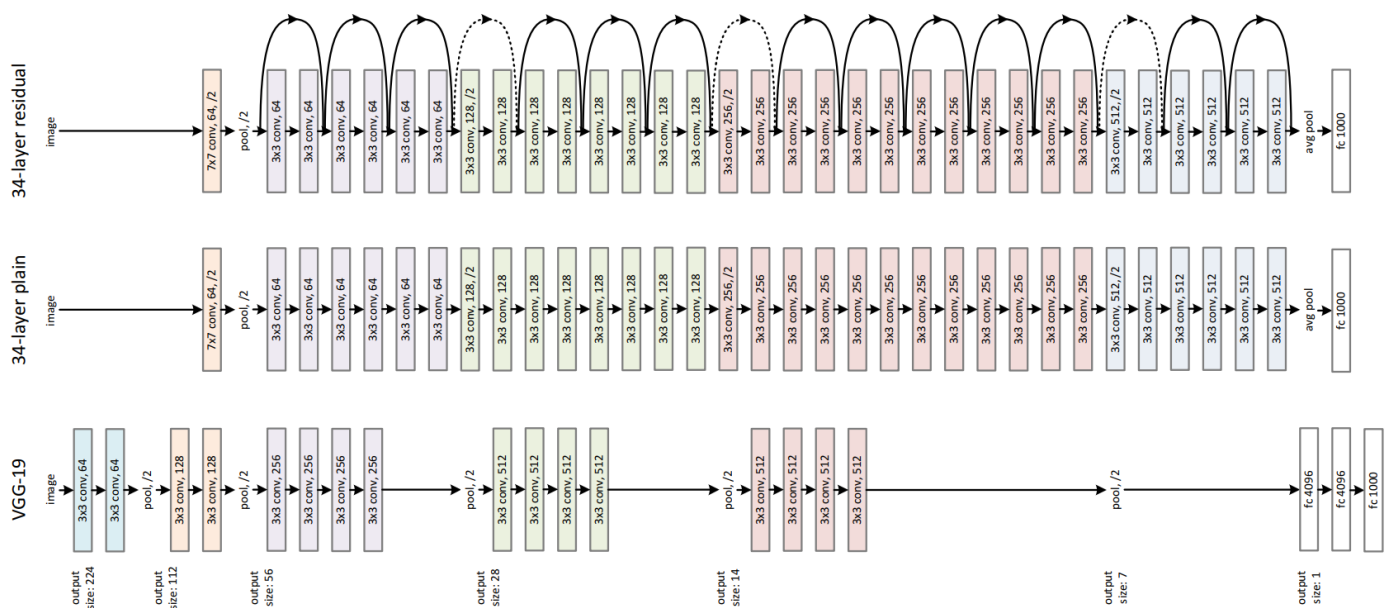
#### 4.2.2.3 ResNet50

Τα ResNet (Residual Neural Networks) ανήκουν στην κατηγορία των «πολύ» Βαθιών Νευρωνικών Δικτύων που, επίσης, διακρίθηκαν για την πρωτοπορία τους. Προτάθηκαν το 2015 από τους Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun [129], δίνοντας λύση στο πρόβλημα αύξησης του ποσοστού σφαλμάτων με αύξηση του βάθους των δικτύων. Πράγματι, υπήρχε η πεποίθηση στον κόσμο της ερευνητικής κοινότητας ότι όσα περισσότερα συνελκτικά στρώματα προστεθούν στο δίκτυο, τόσο καλύτερες θα προκύψουν οι προβλέψεις, δεδομένου του ότι τα επιπλέον επίπεδα θα εκπαιδεύονταν με τέτοιο τρόπο ώστε να συμβάλλουν στην αναγνώριση επιπλέον χαρακτηριστικών και έτσι, συνολικά να επιτευχθούν υψηλότερες επιδόσεις. Ωστόσο, τα αποτελέσματα δεν ήταν τα αναμενόμενα, λόγω του κορεσμού που επέφεραν τα πλεονάζοντα στρώματα, μειώνοντας τα ποσοστά ορθότητας. Αποδείχτηκε ότι αυτό δεν οφειλόταν στην υπερπροσαρμογή (overfitting), αλλά κυρίως στο φαινόμενο των φθινουσών ή εκθετικά αυξανόμενων

κλίσεων (vanishing, exploding gradients). Το πρόβλημα αυτό αντιμετωπίστηκε με τη χρήση των residual blocks.

Ο ιδιαίτερος αυτός τύπος μπλοκ υλοποιεί residual mapping, το οποίο βασίζεται στην αρχή των ‘skip connections’, όπου η έξοδος ενός επιπέδου του δικτύου δεν αποτελεί είσοδο μόνο για το επόμενο από αυτό επίπεδο, αλλά προστίθεται και στην έξοδο κάποιου «μεταγενέστερου» επιπέδου. Αυτή η μέθοδος έχει δύο ευεργετικά οφέλη. Καταρχάς, μετριάζεται το πρόβλημα των φθινουσών κλίσεων, δημιουργώντας εναλλακτικές διαδρομές, ενώ επίσης, με αυτόν τον τρόπο εξασφαλίζεται ότι τα στρώματα που βρίσκονται πιο βαθιά στο δίκτυο δεν ενέχουν κίνδυνο χαμηλότερης επίδοσης σε σχέση με τα αρχικά. Έτσι, με τα ResNet βελτιώθηκε σημαντικά η αποτελεσματικότητα των δικτύων που απαρτιζόνταν από πολλαπλά στρώματα, καθιστώντας ευκολότερη την εκπαίδευσή τους.

Έχουν δημιουργηθεί διαφορετικοί τύποι μοντέλων που ανήκουν σε αυτήν την οικογένεια δικτύων, όπως το ResNet-18, ResNet-34, ResNet-50, ResNet-101, ResNet-110, ResNet-152, ResNet-164 κ.α.

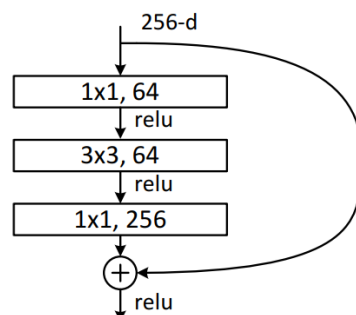


Σχήμα 4.6: Συγκριτική παρουσίαση αρχιτεκτονικών μοντέλων VGG-19 και ResNet. (Πηγή [129])

Στην παρούσα εργασία χρησιμοποιήθηκε το ResNet-50, το οποίο αποτελείται από 49 συνελκτικά επίπεδα, ανάμεσα από τα οποία δεν παρεμβάλλονται συγκεντρωτικά στρώματα, παρά μόνο μετά το πρώτο και το τελευταίο και 1 πλήρως συνδεδεμένο επίπεδο. Η ιδιαιτερότητα του συγκεκριμένου μοντέλου εμφανίζεται στο γεγονός ότι «στοιβάζονται» 3 συνελκτικά επίπεδα (αντί για 2 όπως στο ResNet-34), λόγω των συνδέσεων ‘skip



connections' που δημιουργούνται. Στο επίπεδο εισόδου δέχονται εικόνες διαστάσεων  $224 \times 224 \times 3$ .



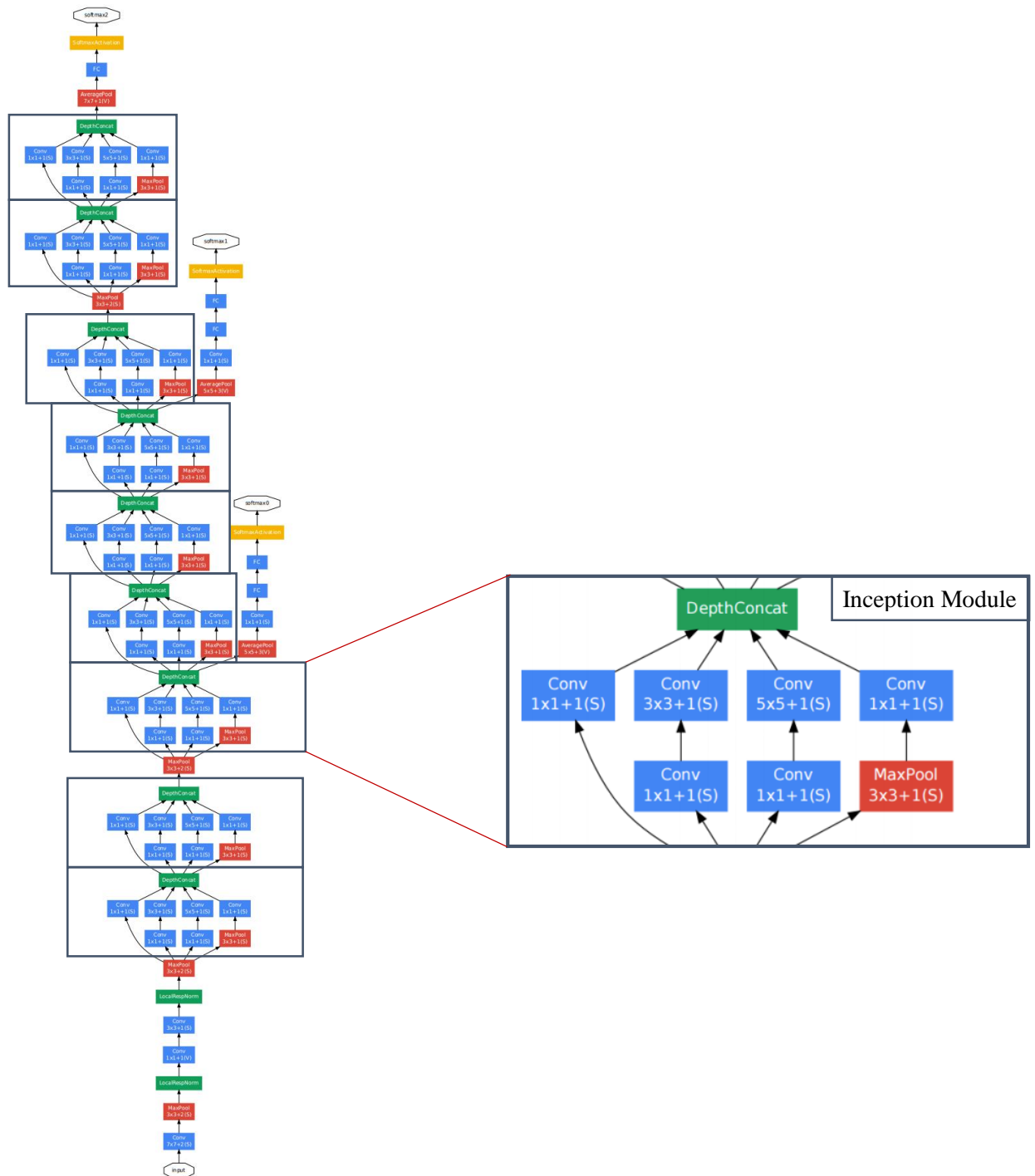
Σχήμα 4.7: Αρχιτεκτονική μοντέλων ResNet50/101/152. (Πηγή [129])

Τέλος, κρίνεται σημαντικό να αναφερθεί ότι τα ResNet αποτέλεσαν μια καινοτομία που βελτίωσε τη διαδικασία εκπαίδευσης των Βαθιών Συνελικτικών Νευρωνικών Δικτύων σε πληθώρα εργασιών συσχετιζόμενες με την όραση υπολογιστών. Κατέκτησαν την πρώτη θέση στο ILSVRC-2015, πετυχαίνοντας εξαιρετικά χαμηλό error rate σε ένα πρόβλημα ταξινόμησης και έκτοτε χρησιμοποιούνται ευρέως.

#### 4.2.2.4 Inception

Το μοντέλο αυτό αποτελεί ορόσημο στην εξέλιξη των Συνελικτικών Νευρωνικών Δικτύων. Πρόκειται για μια αρχιτεκτονική ιδιαίτερα αποδοτική από υπολογιστικής άποψης, τόσο όσον αφορά τον αριθμό των παραμέτρων, όσο και το οικονομικό κόστος. Η πρώτη έρευνα για τα Inception Nets δημοσιεύτηκε το 2014 [130] σε μια προσπάθεια να επιλυθούν προβλήματα που εμφανίζονταν σε διάφορα συνελικτικά μοντέλα και σχετίζονταν με την επιλογή του κατάλληλου μεγέθους πυρήνα, την υπερπροσαρμογή των δικτύων, καθώς και το γεγονός ότι τα βαθιά δίκτυα ήταν και υπολογιστικά ακριβά. Οι συγκεκριμένοι ερευνητές υποστήριξαν ότι ένα δίκτυο πιο «πλατύ» μπορεί να έχει μεγαλύτερη απόδοση σε σχέση με ένα «στενό» βαθύ δίκτυο. Πιο αναλυτικά, πρότειναν τη χρήση πολλαπλών πυρήνων διαφορετικών διαστάσεων που όλοι να λειτουργούν στο ίδιο επίπεδο του εκάστοτε δικτύου. Τα αποτελέσματα που προκύπτουν από τις συνελίξεις της κάθε εισόδου με τα διαφορετικά αυτά φίλτρα, καθώς και η εφαρμογή ενός συγκεντρωτικού τελεστή (pooling operator) στην εν λόγω είσοδο, συγχωνεύονται (concatenated) και καταλήγουν στο επόμενο επίπεδο. Όλα αυτά συγκροτούν μία μονάδα inception (inception module). Το μοντέλο αυτό ονομάστηκε GoogLeNet ή αλλιώς Inception-v1 και περιλαμβάνει 9 τέτοιες μονάδες, ενώ έχει βάθος 22 συνολικά επιπέδων (συνελικτικών και πλήρως συνδεδεμένων).

Σαν συνέχεια στο ερευνητικό τους έργο, παρουσίασαν δύο βελτιωμένα μοντέλα αυτής της οικογένειας, το Inception-v2 και Inception-v3, ενώ αργότερα δημιούργησαν το Inception-v4, καθώς και τα Inception – Resnet v1 και v2. Στην διπλωματική αυτή εργασία χρησιμοποιήθηκε η τρίτη εκδοχή της Inception αρχιτεκτονικής (Inception-v3), η οποία διαφοροποιείται σε σχέση με την αρχική ιδέα στα εξής:



Σχήμα 4.8: Αρχιτεκτονική μοντέλου Inception. (Πηγή [130])

- Αντικατάσταση πυρήνων μεγαλύτερου μεγέθους με διαδοχικούς πυρήνες μικρότερου μεγέθους, για αύξηση της υπολογιστικής ταχύτητας. (Για παράδειγμα, 5×5 φίλτρα αντικαταστάθηκαν το καθένα με δύο 3×3 φίλτρα, μειώνοντας έτσι τον αριθμό των παραμέτρων.)
- Διάσπαση συνελίξεων της μορφής  $n \times n$  σε ένα συνδυασμό  $1 \times n$  και  $n \times 1$  συνελίξεων, καθώς αποδείχτηκε φθηνότερη οικονομικά μέθοδος που αποφέρει τα ίδια αποτελέσματα.
- Αποτελεσματική μείωση των διαστάσεων της εκάστοτε εισόδου, με την αποφυγή της παραδοσιακής τεχνικής, όπου η είσοδος περνά από το συνελικτικό επίπεδο και στη συνέχεια από το συγκεντρωτικό επίπεδο (ή και το αντίστροφο). Τη θέση αυτή της τεχνικής πήρε η μέθοδος κατά την οποία η είσοδος περνά ‘παράλληλα’ από ένα συνελικτικό επίπεδο (που περιλαμβάνει  $1 \times 1$  φίλτρα) και ένα συγκεντρωτικό επίπεδο και τα αποτελέσματα αυτών τελικά συγχωνεύονται. Η μέθοδος αυτή προτιμήθηκε λόγω του μικρότερου οικονομικού κόστους που αποφέρει.

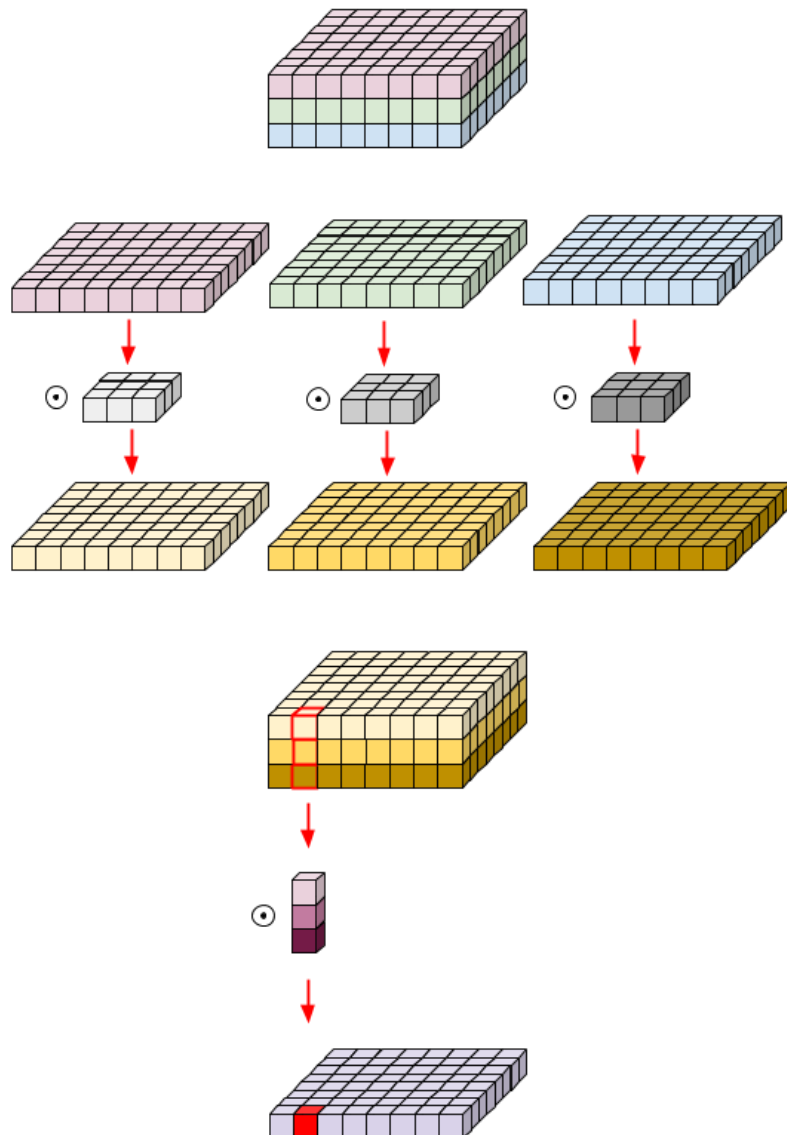
Τα Inception μοντέλα ανέβηκαν, επίσης, σε υψηλές θέσεις στους διαγωνισμούς ILSVRC-2014 και ILSVRC-2015, πετυχαίνοντας αξιοσημείωτες επιδόσεις στα προβλήματα ταξινόμησης, όπου και χρησιμοποιήθηκαν. Σαν είσοδο δέχονται εικόνες διαστάσεων  $299 \times 299 \times 3$ .

#### 4.2.2.5 Xception

Πρόκειται για ένα μοντέλο που, επίσης, ανήκει στην κατηγορία των Βαθιών Νευρωνικών Δικτύων και έχει εμπνευστεί από την αρχιτεκτονική του Inception, γι’ αυτό και ονομάστηκε Xception που αποτελεί ακρωνύμιο του “Extreme Inception”. Δημιουργήθηκε από μια ομάδα ερευνητών της Google και προτάθηκε από τον Francois Chollet το 2016 [131], οπότε αποτελεί το πιο πρόσφατο, εκ των μοντέλων που παρουσιάστηκαν. Η ιδιαιτερότητά του έγκειται στη χρήση των κατά Βάθος Διαχωρίσιμων Συνελίξεων (Depthwise Separable Convolutions), μιας μεθόδου που διαφοροποιείται από τις απλές, κλασικές συνελίξεις, μειώνοντας σημαντικά τον αριθμό των υπολογισμών στο εν λόγω επίπεδο του δικτύου και κατ’ επέκταση των παραμέτρων. Αυτό έχει ως φυσικό επακόλουθο την βελτιωμένη συμπεριφορά αυτού του μοντέλου ως προς το πρόβλημα της υπερπροσαρμογής. Η αρχιτεκτονική Xception στηρίζεται σε δύο βασικά σημεία:

- ✓ στις κατά Βάθος Διαχωρίσιμες Συνελίξεις και
- ✓ στα ‘skip connections’ των ResNet.

Όσον αφορά τον διαφορετικό αυτό τύπο συνελίξεων, προέκυψε από την ιδέα του ότι η διάσταση του βάθους, σε έναν πυρήνα – φίλτρο, μπορεί να διαχωριστεί από τις άλλες δύο διαστάσεις του (πλάτος \* ύψος). Ακριβώς σε αυτή την αρχή βασίζονται οι κατά Βάθος Διαχωρίσιμες Συνελίξεις, οι οποίες για να δώσουν το επιθυμητό τελικό αποτέλεσμα, τους χάρτες χαρακτηριστικών, εκτελούν συνελίξεις σε δύο στάδια: τη συνέλιξη κατά βάθος (Depthwise Convolution) και τη σημειακή συνέλιξη (Pointwise Convolution). Αρχικά, λαμβάνει χώρα η συνέλιξη κατά βάθος, η οποία προϋποθέτει τόσο η εικόνα – είσοδος, όσο και ο πυρήνας – φίλτρο να είναι διαχωρισμένα σε τόσα μέρη όσα είναι τα κανάλια της εικόνας (εάν δηλαδή έχουμε μια έγχρωμη εικόνα με 3 κανάλια, τότε η εικόνα και το φίλτρο «σπάνε» στα 3, όπως φαίνεται παρακάτω, στην εικόνα).

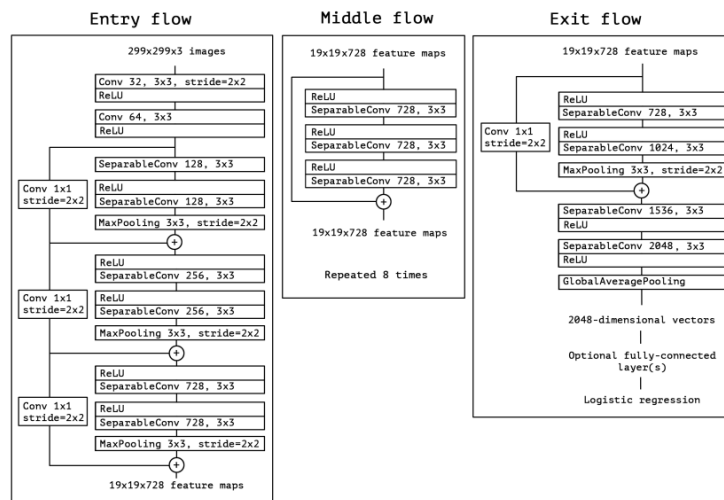


Σχήμα 4.9: Διαδικασία κατά Βάθος Διαχωρίσιμης Συνέλιξης. (Πηγή [132])

Έτσι, πραγματοποιείται συνέλιξη σε κάθε ένα κανάλι της εικόνας-εισόδου ξεχωριστά με ένα δισδιάστατο φίλτρο και σαν αποτέλεσμα προκύπτουν τόσοι πίνακες, όσοι και τα αρχικά κανάλια. Στη συνέχεια, τα αποτελέσματα αυτά στοιβάζονται, δημιουργώντας έναν μόνο πίνακα, ο οποίος συνελίσσεται με έναν πυρήνα διαστάσεων  $1 \times 1 \times C$ , όπου  $C$  το πλήθος των καναλιών. Η διαδικασία αυτή ονομάζεται σημειακή συνέλιξη και δίνει σαν τελικό αποτέλεσμα τον χάρτη χαρακτηριστικών. Με αυτόν τον τρόπο, λοιπόν, καταλήγουν οι κατά βάθος διαχωρίσιμες συνελίξεις στο σημείο όπου καταλήγουν και οι «παραδοσιακές», με τη διαφορά ότι μειώνεται σημαντικά η απαιτούμενη υπολογιστική ισχύς και συνεπώς και το κόστος.

Το Xception χρησιμοποιεί τη μέθοδο αυτή αντί της κλασικής, με μια διαφοροποίηση στη σειρά των συνελίξεων: πρώτα γίνεται Pointwise και μετά Depthwise Convolution. Συνολικά αποτελείται από 36 συνελικτικά επίπεδα, όπου εφαρμόζονται οι κατά Βάθος Διαχωρίσιμες Συνελίξεις. Τα συνελικτικά αυτά επίπεδα στοιβάζονται με τέτοιο τρόπο, ώστε να σχηματίζουν 14 μονάδες, που συνδέονται μέσω των ‘skip connections’.

Ως είσοδο δέχεται εικόνες διαστάσεων  $299 \times 299 \times 3$ . Τα δεδομένα περνούν πρώτα από τη ροή εισόδου (entry flow), στη συνέχεια από τη μεσαία ροή (middle flow), η οποία επαναλαμβάνεται οκτώ φορές, και τέλος από τη ροή εξόδου (exit flow), όπως φαίνεται στο παρακάτω σχήμα.



Σχήμα 4.10: Αρχιτεκτονική μοντέλου Xception. (Πηγή [131])

Το μοντέλο Xception υπερβαίνει τις επιδόσεις των VGG-16, ResNet, ακόμη και Inception-v3 σε αρκετά κλασικά προβλήματα ταξινόμησης. Μάλιστα, στο διαγωνισμό ILSVRC-2016 πέτυχε την 1<sup>η</sup> θέση στην κατηγορία "Image Classification", η οποία απαιτεί την αναγνώριση και την ταξινόμηση εικόνων σε 1.000 διαφορετικές κλάσεις.



## Κεφάλαιο 5

### 5 Αξιολόγηση αποτελεσμάτων

---

Το παρόν κεφάλαιο είναι αφιερωμένο στην παρουσίαση και αξιολόγηση των αποτελεσμάτων, τα οποία προέκυψαν από τα πειράματα και τις εκτενείς δοκιμές που πραγματοποιήθηκαν, έχοντας σαν στόχο τη συσταδοποίηση κινησιογραμμάτων Ελληνικού Παραδοσιακού Χορού. Το κομμάτι αυτό της εργασίας διαδραματίζει σημαντικό ρόλο στην ερευνητική διαδικασία, καθώς επιτρέπει την ανάλυση των δεδομένων και την κατανόηση της συμπεριφοράς των μοντέλων που χρησιμοποιήθηκαν, ενώ ανοίγει τις πύλες για την εξαγωγή χρήσιμων συμπερασμάτων, αλλά και για ενδεχόμενες ερευνητικές επεκτάσεις.

Αρχικά, εξετάζονται διαφορετικές μετρικές που αξιολογούν το πρόβλημα της εν λόγω εργασίας, καθεμία από τις οποίες χρησιμοποιώντας τα δικά της κριτήρια, δίνει διαφορετική «χροιά» όσον αφορά την εγκυρότητα της Συσταδοποίησης.

Στη συνέχεια, εφόσον έχουν γίνει κατανοητές οι μετρικές αξιολόγησης που χρησιμοποιήθηκαν, γίνεται μια συνοπτική περιγραφή σε πρακτικό επίπεδο για κάθε μοντέλο, περί της αρχιτεκτονικής του και του μεγέθους των εξαγόμενων κάθε φορά χαρακτηριστικών. Επιπλέον, γίνεται αναφορά στην τεχνική της σωρευτικής επεξηγούμενης διακύμανσης (*cumulative explained variance*), η οποία χρησιμοποιείται ευρέως σε συνδυασμό με την ανάλυση κυρίων συνιστωσών – PCA, προκειμένου να συμβάλλει στην καταλληλότερη επιλογή αριθμού components. Πιο συγκεκριμένα, για κάθε αριθμό κυρίων συνιστωσών, μετρά το ποσοστό της συνολικής διακύμανσης που αντιπροσωπεύει, συμπεριλαμβάνοντας τον ίδιο και όλους τους προηγούμενους του. Με απλά λόγια, για κάθε αριθμό κυρίων συνιστωσών μας παρέχει το ποσοστό της πληροφορίας που συγκρατεί σε σχέση με την αρχικό όγκο πληροφορίας των δεδομένων. Για το λόγο αυτό, δίνεται και το αντίστοιχο διάγραμμα που μας βοήθησε στη διαδικασία επιλογής, συνδυαστικά με τις δοκιμές που έγιναν, συγκρίνοντας τις τελικές αποδόσεις των μοντέλων κάθε φορά.

Επόμενο βήμα υπήρξε η εφαρμογή του αλγορίθμου συσταδοποίησης. Εδώ επιλέχθηκε ο k-means, μια ευρέως χρησιμοποιούμενη μέθοδος σε προβλήματα clustering, λόγω της απλότητας, της εύκολης προσαρμογής του σε νέα δεδομένα, της ικανότητας γενίκευσης,

καθώς και της κλιμακωσιμότητάς του. Όπως έχει αναλυθεί στο δεύτερο κεφάλαιο της εργασίας, ο αλγόριθμος αυτός έχει σαν στόχο την ομαδοποίηση των δεδομένων, δηλαδή των χαρακτηριστικών στα οποία έχουμε, τελικά, καταλήξει, σε  $k$  προκαθορισμένες συστάδες. Ως αριθμό συστάδων, λοιπόν, επιλέξαμε  $num\_clusters = 21$ , εφόσον στην έρευνά μας κάνουμε λόγο για 21 διαφορετικά είδη χορού και με βάση τον αριθμό αυτό γίνεται ο διαχωρισμός τους σε ομάδες και παρουσιάζονται τα τελικά αποτελέσματα της εργασίας. Ωστόσο, να σημειωθεί ότι έγιναν πειραματισμοί με μικρότερο, αλλά και μεγαλύτερο αριθμό συστάδων, ώστε να παρατηρήσουμε τη συμπεριφορά του αλγορίθμου, με τα συμπεράσματα αυτά να σχολιάζονται παρακάτω. Για την οπτικοποίηση των αποτελεσμάτων που προκύπτουν από το κάθε μοντέλο παρατίθεται ένα διάγραμμα που απεικονίζει την ποσοστιαία αναλογία κάθε είδους χορού για 6 clusters, έτσι ώστε να γίνει περισσότερο κατανοητή η συμπεριφορά του κάθε μοντέλου ως προς το σύνολο δεδομένων.

Τέλος, παρουσιάζονται και σχολιάζονται τα συγκεντρωτικά αποτελέσματα που παράγονται από τα οκτώ διαφορετικά μοντέλα νευρωνικών δικτύων που χρησιμοποιήθηκαν, ώστε να γίνει μια εποικοδομητική σύγκριση και προσπάθεια για ερμηνεία τους.

## 5.1 Μετρικές Αξιολόγησης Συσταδοποίησης

Παρακάτω παρουσιάζονται διαδεδομένες μετρικές για την αξιολόγηση της συσταδοποίησης. Είναι σημαντικό κάθε φορά να γίνεται σωστή επιλογή των μετρικών που χρησιμοποιούνται ανάλογα με τον τύπο του προβλήματος και τη φύση του συνόλου δεδομένων.

### 5.1.1 NMI score

Το NMI (Normalized Mutual Information) ανήκει στην οικογένεια των Mutual Information based scores και μετρά τη συμφωνία μεταξύ των πραγματικών ετικετών των δειγμάτων (ground truth labels) και εκείνων που προβλέπονται από τον αλγόριθμο (predicted labels). Παίρνει τιμές μεταξύ του 0 και 1, με το 1 να αναπαριστά το υψηλότερο σκορ για εντελώς σωστή ομαδοποίηση.

Για τον υπολογισμό του χρησιμοποιεί την έννοια της εντροπίας (entropy), μια μετρική που ποσοτικοποιεί την αβεβαιότητα και ορίζεται ως εξής:



$$H(T) = - \sum_{i=1}^t P(T_i) \log(P(T_i))$$

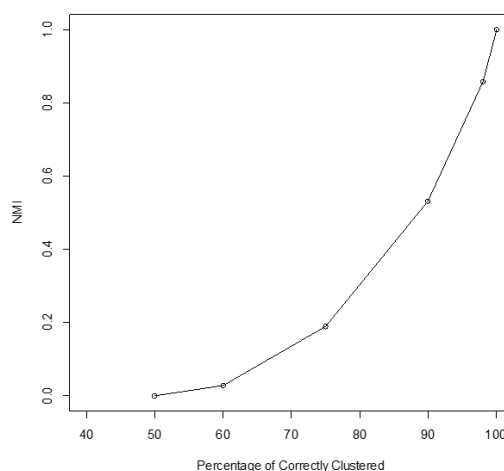
όπου  $t$  είναι ο αριθμός των πραγματικών κλάσεων και  $T_i$  είναι η πραγματική κλάση  $i$ .

Θα μπορούσαμε να πούμε ότι το NMI μας δείχνει το πόσο μειώνεται η αβεβαιότητα σχετικά με τις κλάσεις των δειγμάτων, όταν γνωρίζουμε τις ετικέτες των συστάδων. Η εξίσωση που χρησιμοποιείται για τον υπολογισμό του:

$$\eta = \frac{2(H(T) - H(T|C))}{H(T) + H(C)}$$

όπου  $C$  θεωρείται το σύνολο των συστάδων που έχει προκύψει σαν αποτέλεσμα της συσταδοποίησης.

Να σημειωθεί ότι πρόκειται για μια μη γραμμική μετρική αξιολόγησης της επίδοσης του clustering. Αυτό έχει σαν αποτέλεσμα, όταν, για παράδειγμα, τα μισά δείγματα από το σύνολο δεδομένων ομαδοποιούνται σωστά να προκύπτει τιμή κοντά στο 0, σε αντίθεση με κάποια γραμμική μετρική, όπου θα είχαμε σκορ κοντά στο 0.5. Αντιστοίχως, όπως φαίνεται και στην παρακάτω εικόνα, NMI score = 0.5 όταν περίπου το 88% των παρατηρήσεων έχουν ομαδοποιηθεί σωστά και NMI score = 0.6 όταν το 93% αυτών ανήκουν στις σωστές συστάδες [133]. Πρόκειται, λοιπόν, για μια «αυστηρή» μετρική αξιολόγησης που εν γένει θεωρείται αξιόπιστη και έχει χρησιμοποιηθεί συχνά στη βιβλιογραφία.



Σχήμα 5.1: Συσχέτιση του NMI score με το ποσοστό των παρατηρήσεων που έχουν ομαδοποιηθεί σωστά.

(Πηγή [133])

### 5.1.2 Rand index score

Πρόκειται για ένα κριτήριο αξιολόγησης, το οποίο μετρά την ομοιότητα μεταξύ των ground truth labels και predicted labels. Παίρνει τιμές μεταξύ του 0 και 1, και όσο μεγαλύτερο είναι το σκορ τόσο καλύτερη και η συσταδοποίηση.

Για τον υπολογισμό του χρησιμοποιεί ζευγάρια δειγμάτων, τα οποία προκύπτουν συνδυάζοντας το σύνολο των παρατηρήσεων ανά δύο. Στη συνέχεια, τα ζεύγη διαχωρίζονται σε 4 κατηγορίες:

- TP – True Positives  
Τα δύο δείγματα έχουν την ίδια ground truth και την ίδια predicted ετικέτα.
- TN – True Negatives  
Τα δύο δείγματα έχουν διαφορετική ground truth και διαφορετική predicted ετικέτα.
- FP – False Positives  
Τα δύο δείγματα έχουν διαφορετική ground truth, αλλά την ίδια predicted ετικέτα.
- FN – False Negatives  
Τα δύο δείγματα έχουν διαφορετική ground truth, αλλά την ίδια predicted ετικέτα.

Ουσιαστικά, η μετρική αυτή υπολογίζει το ποσοστό των ζευγών που έχουν ομαδοποιηθεί “σωστά”, ως εξής:

$$RI = \frac{TP + TN}{TP + TN + FP + FN}$$

Η μετρική αυτή είναι ευρέως χρησιμοποιούμενη σε προβλήματα συσταδοποίησης, ιδίως για τη σύγκριση διαφορετικών αλγορίθμων επί του ίδιου συνόλου δεδομένων. Παρουσιάζει, ωστόσο, το μειονέκτημα του ότι τα FP και FN έχουν τον ίδιο συντελεστή βαρύτητας, κάτι το οποίο μπορεί να επηρεάσει το αποτέλεσμα, καθώς κάποιες φορές το να διαχωριστούν δείγματα που ανήκουν στην ίδια κλάση είναι χειρότερο από το να ομαδοποιηθούν δείγματα που ανήκουν σε διαφορετικές [134].

### 5.1.3 Davies-Bouldin Index

Η μετρική αυτή αντιπροσωπεύει τη μέση ομοιότητα της κάθε συστάδας με την περισσότερη όμοια συστάδα προς αυτή [73]. Η μικρότερη τιμή που μπορεί να λάβει είναι

το 0, ωστόσο δεν υπάρχει όριο για τη μέγιστη τιμή. Σε αντίθεση με τις υπόλοιπες μετρικές που έχουν παρουσιαστεί, όσο μικρότερη είναι η τιμή της, τόσο εγκυρότερη η συσταδοποίηση. Αυτό εξηγείται, δεδομένου του ότι είναι επιθυμητό για τις συστάδες να παρουσιάζουν τη μικρότερη δυνατή ομοιότητα μεταξύ τους, οπότε αναλόγως μειώνεται και ο δείκτης Davies-Bouldin.

Όσον αφορά το μαθηματικό τύπο που χρησιμοποιείται για τον υπολογισμό του:

$$DB = \frac{1}{n} \sum_{i=1}^n \max_{i \neq j} \frac{s_i + s_j}{d_{ij}}$$

όπου  $n$  είναι ο αριθμός των συστάδων

$s_i$  είναι μέτρο ομοιότητας εντός της συστάδας  $i$ ,

$s_j$  είναι μέτρο ομοιότητας εντός της συστάδας  $j$  και

$d_{ij}$  είναι η απόσταση μεταξύ των κέντρων των συστάδων  $i$  και  $j$ .

Χαμηλές τιμές αυτού του κριτηρίου είναι ενδεικτικές της ύπαρξης συμπαγών και καλώς διαχωρισμένων ομάδων.

Πρόκειται για μια ακόμη αξιόπιστη και συχνά χρησιμοποιούμενη μετρική σε προβλήματα συσταδοποίησης, η οποία συνδυαζόμενη με άλλες μπορεί να δώσει μια καλύτερη κατανόηση σχετικά με την εγκυρότητα της διαδικασίας.

#### 5.1.4 Purity score

Πρόκειται για μια απλή μετρική αξιολόγησης που χρησιμοποιεί τις πραγματικές ετικέτες των παρατηρήσεων (ground truth labels). Αντιπροσωπεύει το ποσοστό των δειγμάτων εντός μιας συστάδας, τα οποία ανήκουν στην κλάση με την υψηλότερη συχνότητα εμφάνισης στην εν λόγω συστάδα. Λαμβάνει τιμές μεταξύ 0 και 1, με το 0 να αναπαριστά μια αποτυχημένη συσταδοποίηση και το 1 μια απολύτως επιτυχημένη.

Πιο συγκεκριμένα, για τον υπολογισμό αυτής της μετρικής, το κάθε cluster αντιστοιχίζεται, αρχικά, με την κατηγορία στην οποία ανήκει η πλειοψηφία των δειγμάτων που περιλαμβάνει, οπότε στην κάθε συστάδα δίνεται ένα όνομα – ετικέτα. Στη συνέχεια, για κάθε cluster υπολογίζεται ο αριθμός των δειγμάτων του που συμφωνούν με την «επικρατούσα» κλάση. Ακολουθεί πρόσθεση των αριθμών αυτών, που προκύπτουν από κάθε cluster, και στη συνέχεια διαίρεση με το συνολικό αριθμό των δειγμάτων.

$$purity(\Omega, C) = \frac{1}{N} \sum_k \max_j |\omega_k \cap c_j|$$

όπου  $N$  είναι ο αριθμός των δειγμάτων του συνόλου δεδομένων,

$\Omega = \{\omega_1, \omega_2, \dots, \omega_k\}$  είναι οι συστάδες και

$C = \{c_1, c_2, \dots, c_j\}$  είναι οι κλάσεις.

Να σημειωθεί, ωστόσο, ότι αυτή η μετρική αυξάνεται, εν γένει, με την αύξηση του αριθμού των συστάδων, κάτι το οποίο θα πρέπει να λαμβάνεται σοβαρά υπόψη όταν αξιολογείται η ποιότητα της εκάστοτε συσταδοποίησης.

Φυσικά, υπάρχουν ακόμη αρκετές μετρικές που μπορούν να χρησιμοποιηθούν σε τέτοιου είδους προβλήματα. Ωστόσο, δε θα προβούμε σε ανάλυσή τους, καθώς δε χρησιμοποιούνται στα πλαίσια της εν λόγω διπλωματικής εργασίας.

## 5.2 Αποτελέσματα μη προ-εκπαιδευμένων μοντέλων

Σε πρώτη φάση, γίνεται παρουσίαση των αποτελεσμάτων, τα οποία προέκυψαν από τα τρία συνελκτικά νευρωνικά δίκτυα που δημιουργήθηκαν για το συγκεκριμένο πρόβλημα συσταδοποίησης και δεν έχουν προ-εκπαιδευτεί σε προηγούμενο σύνολο δεδομένων.

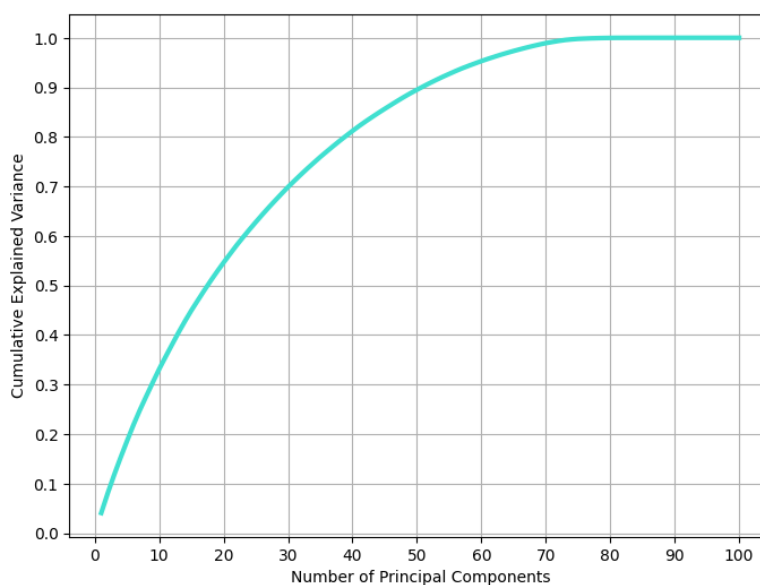
### 5.2.1 Μοντέλο Πρώτο

Όπως αναλύθηκε στο προηγούμενο κεφάλαιο, πρόκειται για ένα εξαιρετικά απλής αρχιτεκτονικής δίκτυο, το οποίο μας ενδιαφέρει για πειραματικούς και μόνο σκοπούς, ώστε να δούμε ποιο είναι, κατά προσέγγιση, το κατώφλι, η απόδοση ενός εντελώς λιτού δικτύου, το οποίο μπορεί να χρησιμοποιηθεί σε σημείο σύγκρισης για τα επόμενα πιο περίπλοκα μοντέλα. Αποτελείται από 3 μόνο layers: 1 convolutional, 1 max pooling, 1 fully connected. Το δίκτυο αυτό εξάγει τα χρήσιμα χαρακτηριστικά του συνόλου δεδομένων, και έτσι προκύπτουν διανύσματα μεγέθους  $1 \times 128$  για κάθε μία από τις 271 εικόνες.

Στη συνέχεια, τα διανύσματα αυτά υφίστανται μείωση διαστατικότητας μέσω της μεθόδου PCA, όπως έχει προαναφερθεί. Πρακτικά, αυτό που έχει σημασία είναι να βρεθεί ο σωστός αριθμός συνιστωσών, εκείνος που είναι όσο το δυνατόν μικρότερος, αποδίδοντας συγχρόνως τη μεγαλύτερη δυνατή ποσότητα πληροφορίας. Φυσικά, προκειμένου να

επιτευχθεί αυτό έγιναν πολλαπλές δοκιμές, ενώ βοηθητικά χρησιμοποιήθηκαν οι τιμές της *cumulative explained variance*. Στόχος είναι να βρεθεί η ελάχιστη τιμή για τα principal components που σταθεροποιεί τη *cumulative explained variance* κοντά στο 1. Προφανώς, όσο μικρότερη είναι αυτή η τιμή, τόσο μεγαλύτερη εξοικονόμηση χρόνου και μνήμης γίνεται στη μετέπειτα διαδικασία του clustering.

Όπως φαίνεται στο παρακάτω διάγραμμα για το πρώτο μοντέλο, μια τιμή πάνω από 60 είναι ιδανική για τον αριθμό των κυρίων συνιστωσών, καθώς ξεπερνά το 95% της συνολικής πληροφορίας. Έτσι, επιλέχθηκε *num\_components* = 65.

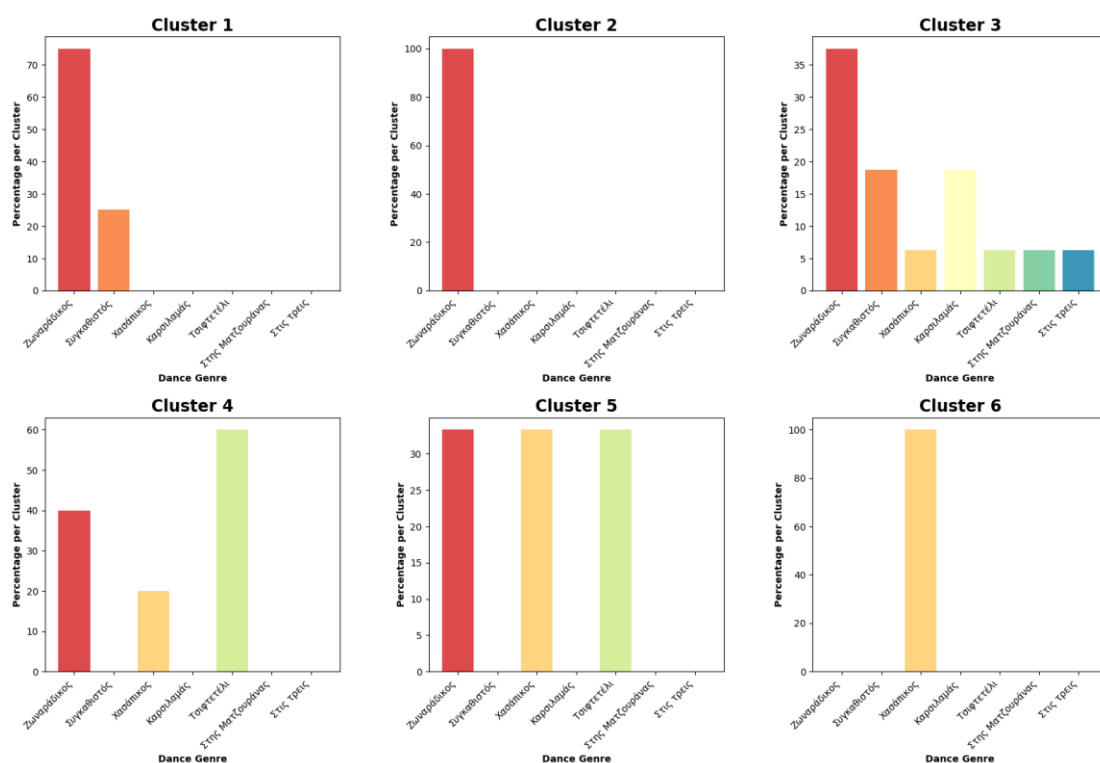


Σχήμα 5.2: Διάγραμμα *cumulative explained variance* για το πρώτο CNN μοντέλο.

Αφού μειώθηκαν οι διαστάσεις των features, διατηρώντας την απαραίτητη πληροφορία, το επόμενο βήμα ήταν η εφαρμογή του αλγορίθμου συσταδοποίησης, k-means που προέρχεται από τη βιβλιοθήκη του scikit-learn. Οι περισσότερες δοκιμές έγιναν όσον αφορά τη ρύθμιση (fine-tuning) δύο υπερπαραμέτρων του. Παρατηρήσαμε ότι οι συγκεκριμένες υπερπαραμέτροι παίζουν σημαντικό ρόλο στα τελικά αποτελέσματα, όπου είδαμε ποσοστιαία διαφορά ίση με 5-10%, ανάλογα με τη μετρική αξιολόγησης κάθε φορά, η οποία βάσει των τιμών που προκύπτουν για το πρόβλημά μας δε θα μπορούσε σε καμία περίπτωση να χαρακτηριστεί αμελητέα. Έτσι, δοκιμάσαμε διαφορετικές τιμές για το *n\_init*, που σχετίζεται με το πλήθος των φορών που θα εκτελεστεί ο αλγόριθμος χρησιμοποιώντας κάθε φορά διαφορετικά αρχικά κέντρα, καθώς και για το *random\_state*, που καθορίζει την παραγωγή ενός αριθμού σχετικά με την αρχικοποίηση των κέντρων.

Τα αποτελέσματα που παρουσιάζονται είναι εκείνα που είχαν τα υψηλότερα ποσοστά επιτυχίας βάσει των μετρικών αξιολόγησης που χρησιμοποιήσαμε. Επειδή, όπως έχει ήδη

εξηγηθεί, ωστόσο, οι μετρικές αυτές εισάγουν «υποκειμενικά» κριτήρια ως προς την αξιολόγηση του προβλήματος, στη συνέχεια, θεωρήσαμε αναγκαίο να εξετάσουμε οι ίδιοι τα όσα παράγει το σύστημα και ακόμη καλύτερα να τα οπτικοποιήσουμε, ώστε να είναι ευανάγνωστα και κατανοητά για τον αναγνώστη, επίσης. Έτσι, παρακάτω παρατίθεται ένα διάγραμμα, όπου απεικονίζονται 6 clusters με τις κατηγορίες χορών που περιλαμβάνουν, τα οποία παράχθηκαν από το πρώτο συνελκτικό μοντέλο της εργασίας. Πιο συγκεκριμένα, κάθε bar plot δείχνει την ποσοστιαία αναλογία κάθε είδους χορού στο συγκεκριμένο cluster.



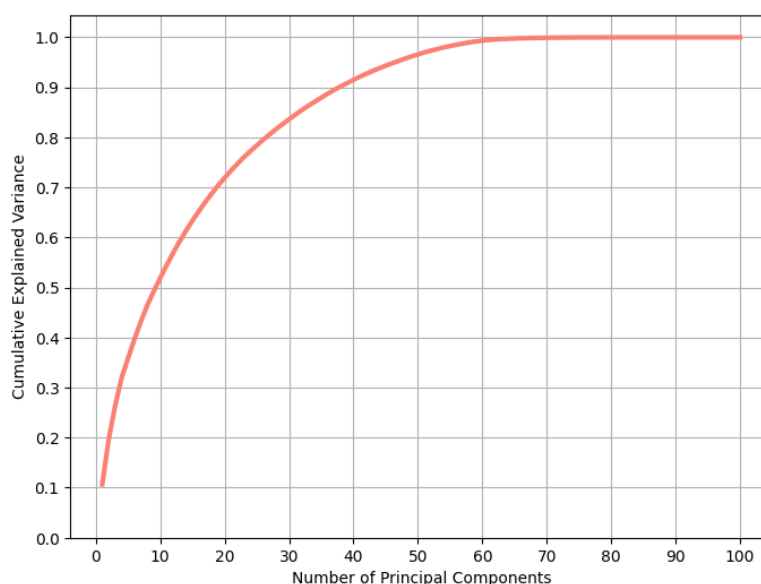
Σχήμα 5.3: Αποτελέσματα clustering για το πρώτο CNN μοντέλο.

Τα αποτελέσματα αυτά, εκ πρώτης όψης φαίνονται σχετικά καλά, δεδομένου του ότι προέρχονται από ένα τόσο απλό συνελκτικό δίκτυο. Παρόλα αυτά, θα ήθελα να επισημάνω ότι εδώ παρουσιάζονται τα ποσοστά εμφάνισης του κάθε είδους χορού, οπότε δεν υπάρχει εικόνα για το πως ακριβώς ομαδοποιήθηκαν τα κινησιογράμματα. Για το συγκεκριμένο μοντέλο, λοιπόν, στο 2<sup>ο</sup> και 6<sup>ο</sup> cluster που αγγίζουν το 100% σε ένα συγκριμένο είδος χορού, έχουν αντιστοιχηθεί στην πραγματικότητα από ένα μόνο κινησιογράμματα (από τα 271 συνολικά), κάτι το οποίο μας δείχνει ότι θα υπάρχει μεγάλη συγκέντρωση παρτιτούρων στις υπόλοιπες συστάδες και το οποίο δεν είναι θεμιτό. Βέβαια, αναμέναμε κάτι τέτοιο εφόσον μιλάμε για ένα εντελώς λιτό αρχιτεκτονικά μοντέλο που δεν έχει προ-εκπαιδευτεί, όχι απλώς σε τέτοιου είδους δεδομένα, αλλά και σε άλλες εικόνες γενικότερα.

## 5.2.2 Μοντέλο Δεύτερο

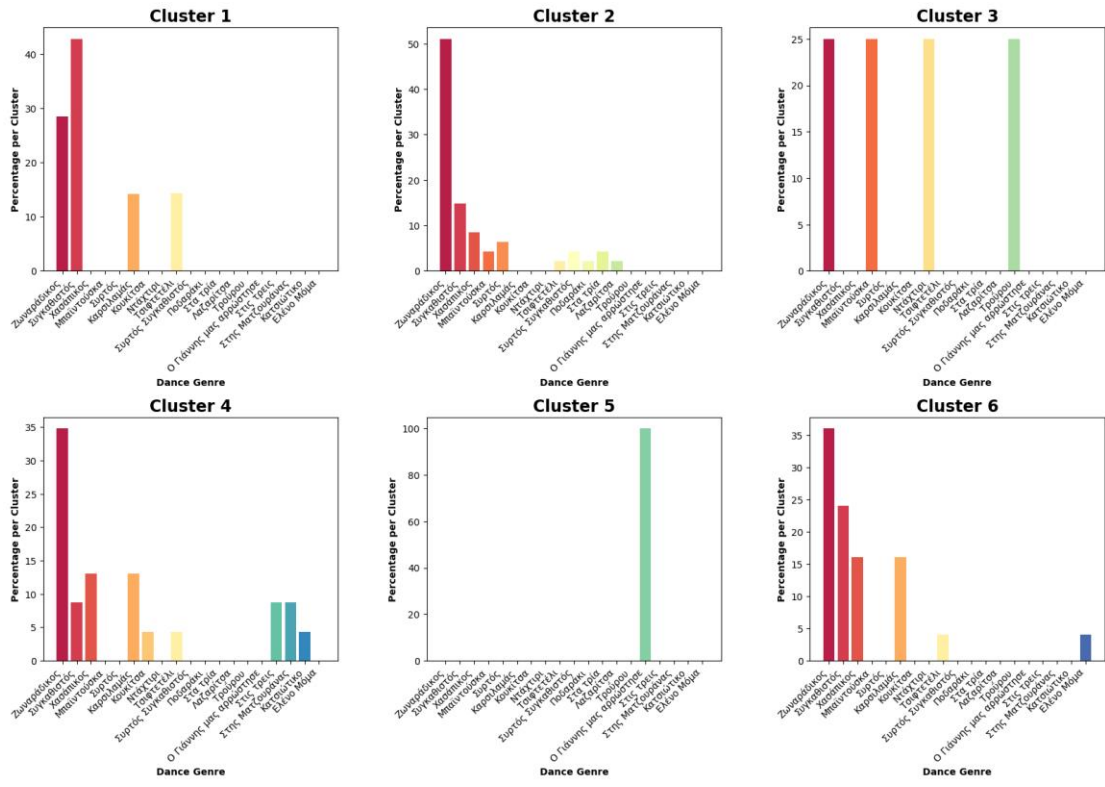
Προχωρώντας στο επόμενο δίκτυο που μελετήθηκε, πρόκειται για ένα μοντέλο, επίσης, απλής αρχιτεκτονικής, μια «αναβάθμιση» του προηγούμενου που περιλαμβάνει τρία πανομοιότυπα convolutional blocks, ώστε να εξεταστεί εάν αυτή η «στοίβαξη» συνελκτικών στρωμάτων βοηθά στη βελτίωση των τελικών αποτελεσμάτων. Επίσης, σε αυτό το μοντέλο χρησιμοποιήθηκε σε κάθε μπλοκ από ένα batch normalization layer που, πράγματι, βελτίωσε σε μεγάλο βαθμό την απόδοση του δικτύου. Πιο συγκεκριμένα, περιλαμβάνει 3 συνελκτικά μπλοκ, καθένα από τα οποία αποτελείται από: 1 convolutional, 1 batch normalization και 1 max pooling layer, ενώ στο τέλος προστίθεται και 1 fully connected στρώμα. Ο αριθμός των εξαγόμενων χαρακτηριστικών είναι κοινός με το πρώτο μοντέλο, οπότε προκύπτουν διανύσματα μεγέθους  $1 \times 128$  για κάθε μία από τις 271 εικόνες. Ακολουθεί η τεχνική PCA, αντιστοίχως με παραπάνω και σε αυτήν την περίπτωση το διάγραμμα της *cumulative explained variance* φαίνεται παρακάτω.

Εδώ, ο αριθμός των κύριων συνιστωσών που απαιτούνται για να διατηρήσουν το μεγαλύτερο μέρος της πληροφορίας είναι ακόμη μικρότερος και κυμαίνεται μεταξύ των τιμών 50 και 60. Για το δεύτερο μοντέλο, λοιπόν, επιλέξαμε *num\_components* = 46.



Σχήμα 5.4: Διάγραμμα *cumulative explained variance* για το δεύτερο CNN μοντέλο.

Στη συνέχεια, παρουσιάζονται τα αποτελέσματα του clustering, δίνοντας μια εικόνα για τις κατηγορίες των χωρών και, πιο συγκεκριμένα, τα ποσοστά αυτών σε κάθε μία από τις πρώτες 6 συστάδες του δεύτερου συνελκτικού μοντέλου.



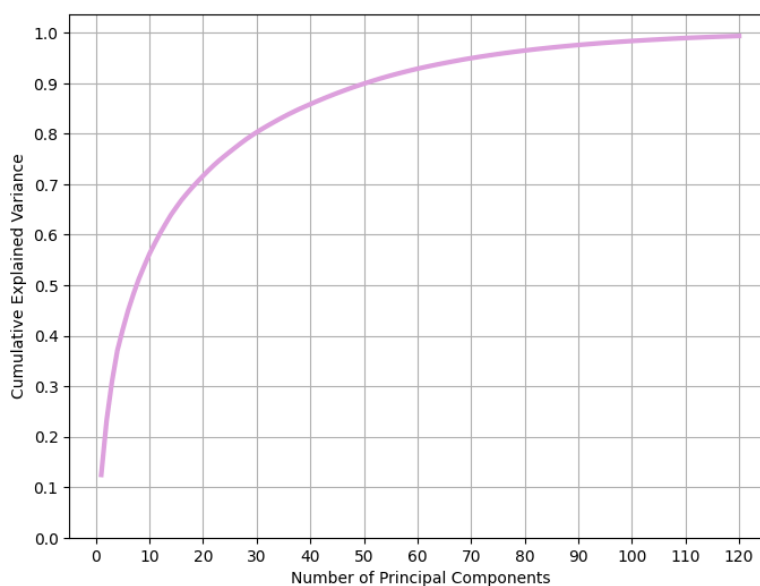
Σχήμα 5.5: Αποτελέσματα clustering για το δεύτερο CNN μοντέλο.

Βάσει του παραπάνω διαγράμματος, συνολικά, παρατηρούμε ότι εν συγκρίσει με το πρώτο μοντέλο, εμφανίζονται πολύ περισσότερες κατηγορίες χορών στα 6 αυτά clusters, το οποίο, βέβαια, μπορεί να είναι και τυχαίο, γιατί εάν διαλέγαμε κάποια άλλα clusters, από τα 21 στο σύνολο, ενδέχεται να μην είχαμε τόσο μεγάλη ποικιλία. Πιο συγκεκριμένα, στο 3<sup>ο</sup> cluster, οι κατηγορίες χορών μοιράζονται τις ίδιες τιμές και, μάλιστα, εάν κάποιος δει αναλυτικά τα κινησιογράμματα που περιλαμβάνονται σε κάθε συστάδα, θα ανακαλύψει ότι αυτή η συστάδα περιλαμβάνει μόνο ένα κινησιογράμματα από κάθε είδος χορού, το οποίο δεν είναι και ιδιαίτερα θετικό. Παρόλα αυτά, στη 2<sup>η</sup>, την 4<sup>η</sup> και την 6<sup>η</sup> ομάδα, όπου είναι προφανές ότι ο Ζωνναράδικος υπερτερεί σε σχέση με τις υπόλοιπες κατηγορίες, έχουμε, πράγματι, αρκετά μεγάλη συγκέντρωση κινησιογραμμάτων στην καθεμία (στη 2<sup>η</sup> συστάδα περιέχονται 47 κινησιογράμματα), κάτι το οποίο είναι θετικό, εφόσον θα μπορούσαμε να πούμε ότι το clustering πέτυχε σε ένα βαθμό. Όσον αφορά, την 5<sup>η</sup> συστάδα που αγγίζει το 100%, επίσης, παρατηρούμε βελτίωση σε σχέση με το πρώτο μοντέλο, καθώς εδώ περιλαμβάνονται 5 παρτιτούρες που, μάλιστα, όχι απλώς ανήκουν στο ίδιο είδος, αλλά αποτελούν παραλλαγές του ίδιου χορού.



### 5.2.3 Μοντέλο Τρίτο

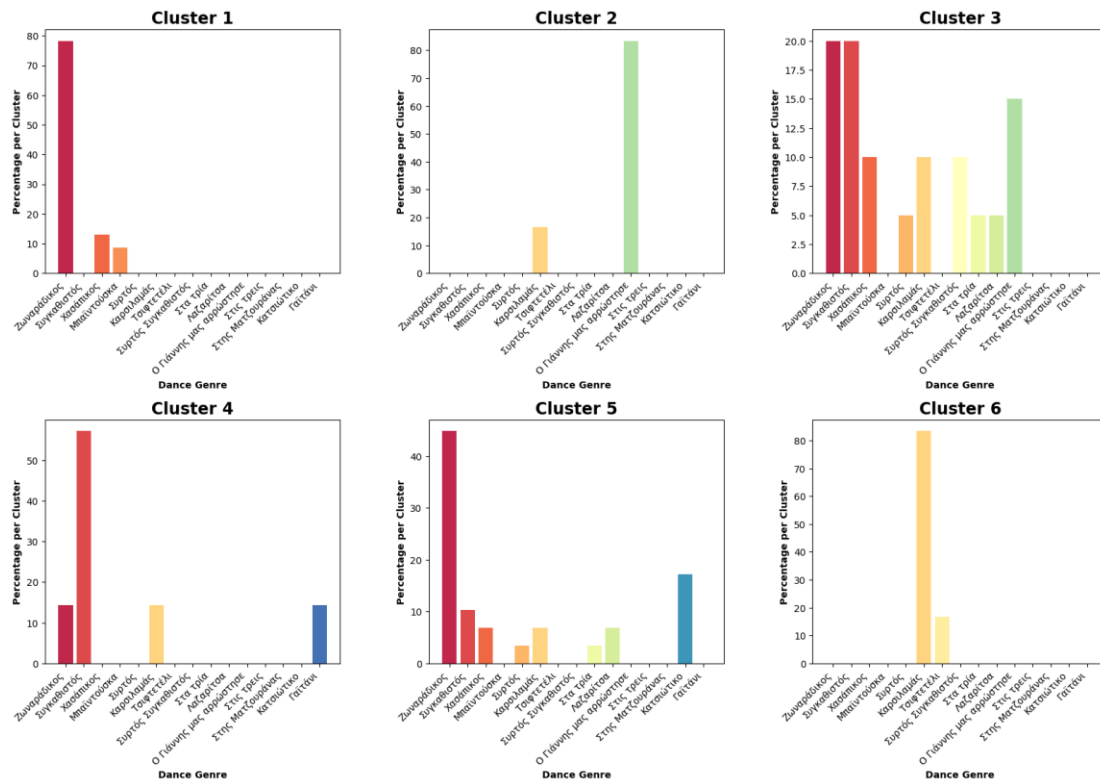
Το τελευταίο από τα from scratch μοντέλα μας, αποτελεί την πολυπλοκότερη εκ των τριών εκδοχών, που στην πραγματικότητα αποτελεί μια ακόμη απλή σχετικά εκδοχή νευρωνικών δικτύων, εάν συγκριθεί με τα μοντέλα που έπονται. Το εν λόγω δίκτυο αποτελείται από τέσσερα συνολικά συνελκτικά μπλοκ, τα οποία είναι ανά δύο πανομοιότυπα. Η διαφορά τους έγκειται στο ότι κάποια χρησιμοποιούν απλά συνελκτικά στρώματα, ενώ τα άλλα χρησιμοποιούν διαχωρίσιμα συνελκτικά στρώματα. Και στις δύο περιπτώσεις καθένα από τα block αποτελείται από: 1 convolutional (είτε απλό είτε separable), 1 batch normalization και 1 max pooling layer, ενώ στο τέλος της δομής προστίθενται 2 fully connected layers, με διαφορετικό αριθμό κόμβων μεταξύ τους. Τελικά, το μοντέλο αυτό που χρησιμοποιήθηκε σαν εξαγωγέας χαρακτηριστικών για τα κνησιογράμματα αποδίδει 271 διανύσματα μεγέθους  $1 \times 256$  το καθένα. Όσον αφορά το διάγραμμα της *cumulative explained variance*, φαίνεται ότι λόγω του μεγαλύτερου αριθμού εξαγόμενων χαρακτηριστικών, η καμπύλη συγκλίνει στο 1 για επίσης υψηλότερο αριθμό principal components.



Σχήμα 5.6: Διάγραμμα *cumulative explained variance* για το τρίτο CNN μοντέλο.

Πιο αναλυτικά, παρατηρούμε ότι για να πετύχουμε το 95% της αρχικής πληροφορίας θα πρέπει να επιλέξουμε αριθμό μεγαλύτερο του 70, ενώ περίπου στο 120 φαίνεται να προσεγγίζει το 1.0. Για αυτό το μοντέλο θέσαμε *num\_components = 76*.

Μετά την εφαρμογή της μεθόδου PCA, ακολουθεί ο αλγόριθμος k-means, τα αποτελέσματα του οποίου, για τα 6 clusters, φαίνονται στο παρακάτω σχήμα.



Σχήμα 5.7: Αποτελέσματα clustering για το τρίτο CNN μοντέλο.

Σύμφωνα με το διάγραμμα, στην 1<sup>η</sup> και 5<sup>η</sup> συστάδα, με 23 και 29 κινησιογράμματα αντίστοιχα, ο Ζωναράδικος κυριαρχεί ενώ, μάλιστα, στο πρώτο εξ αυτών αγγίζει το 78%, κάτι που αποδεικνύει ότι υπάρχει μεγάλη συγκέντρωση παρτιτούρων που απεικονίζουν τα βήματα του χορού αυτού στα συγκεκριμένα clusters. Στην 3<sup>η</sup> συστάδα, ωστόσο, φαίνεται ότι υπάρχουν πολλά είδη χορού, χωρίς να υπερτερεί κάποιο, το οποίο δεν είναι ιδιαίτερος θεμιτό, δεδομένου του ότι ο αλγόριθμος αντιστοίχισε στην ίδια ομάδα αρκετές διαφορετικές κατηγορίες χορών. Φυσικά, λαμβάνοντας υπόψιν το γεγονός ότι πρόκειται για ένα μη προ-εκπαιδευμένο και σχετικά απλό αρχιτεκτονικά μοντέλο, με μικρό βάθος ως προς τα συνελκτικά επίπεδα, είναι λογικό να λαμβάνουμε αυτό το αποτέλεσμα και, σίγουρα, στα υπόλοιπα 15 clusters, θα έχουμε και άλλες τέτοιες συστάδες, με ποικιλία χορευτικών ειδών. Τέλος, όσον αφορά το 6<sup>ο</sup> διάγραμμα, βλέπουμε ότι ο Καρσιλαμάς πετυχαίνει ποσοστό ίσο με 83.33% και, πιο συγκεκριμένα, τυπώνοντας τους τίτλους των χορών που περιλαμβάνει, παρατηρούμε ότι δεν αντιστοιχίζεται ένας μόνο χορός, όπως είχαμε δει στο πρώτο μοντέλο, αλλά 5 χοροί που ανήκουν στο είδος αυτό.

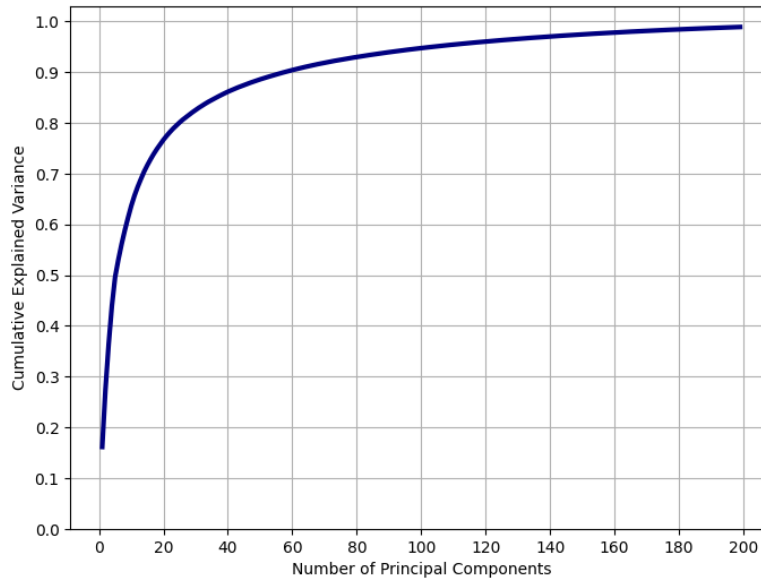
## 5.3 Αποτελέσματα προ-εκπαιδευμένων μοντέλων

Σε δεύτερη φάση, παρουσιάζονται τα αποτελέσματα που προέκυψαν από προ-εκπαιδευμένα συνελκτικά νευρωνικά δίκτυα, ευρέως χρησιμοποιούμενα από την ερευνητική κοινότητα στο πεδίο της μηχανικής μάθησης και όρασης υπολογιστών. Εφόσον έχουν ήδη εκπαιδευτεί σε κάποιο σύνολο δεδομένων, χρησιμοποιούμε την τεχνική της μεταφοράς μάθησης (transfer learning), ώστε να αξιοποιήσουμε την προαποκτηθείσα γνώση τους και να εξετάσουμε κατά πόσο αυτή βελτιώνει τα τελικά αποτελέσματα. Πειραματιστήκαμε συνολικά με πέντε διαφορετικά μοντέλα, ώστε να διερευνήσουμε πόσο αποδοτικά είναι στη συσταδοποίηση κινησιογραμμάτων, αλλά και ποιο από αυτά «ταιριάζει» καλύτερα στο πρόβλημα που μελετά η παρούσα εργασία, ποιο λειτουργεί καλύτερα σαν εξαγωγέας χαρακτηριστικών για τις συγκεκριμένες εικόνες.

### 5.3.1 VGG16

Το πρώτο από τα μοντέλα που εξετάζεται είναι το δημοφιλές VGG16, το οποίο επιλέχθηκε λόγω της ευελιξίας και της αποδοτικότητάς του. Έχει σχετικά μικρό αριθμό παραμέτρων εν συγκρίσει με άλλα δίκτυα, συμβάλλοντας στην εξοικονόμηση χρόνου και υπολογιστικών πόρων. Όσον αφορά την αρχιτεκτονική του, όπως έχει ήδη αναφερθεί αποτελείται από 16 κύρια επίπεδα: 13 convolutional (ανάμεσα στα οποία παρεμβάλλονται και max pooling layers) και 3 fully connected. Σε πρακτικό επίπεδο, εμείς αφαιρέσαμε το τελευταίο πλήρως συνδεδεμένο επίπεδο (Dense layer), αυτό το οποίο ευθύνεται για την ταξινόμηση των εικόνων σε κλάσεις, εφόσον δεν αποτελεί ζητούμενο της εργασίας. Οπότε το ελαφρώς τροποποιημένο δίκτυο που χρησιμοποιήσαμε, εξάγει τα χρήσιμα χαρακτηριστικά του συνόλου δεδομένων, και έτσι προκύπτουν διανύσματα μεγέθους  $1 \times 4096$  για κάθε μία από τις 271 εικόνες.

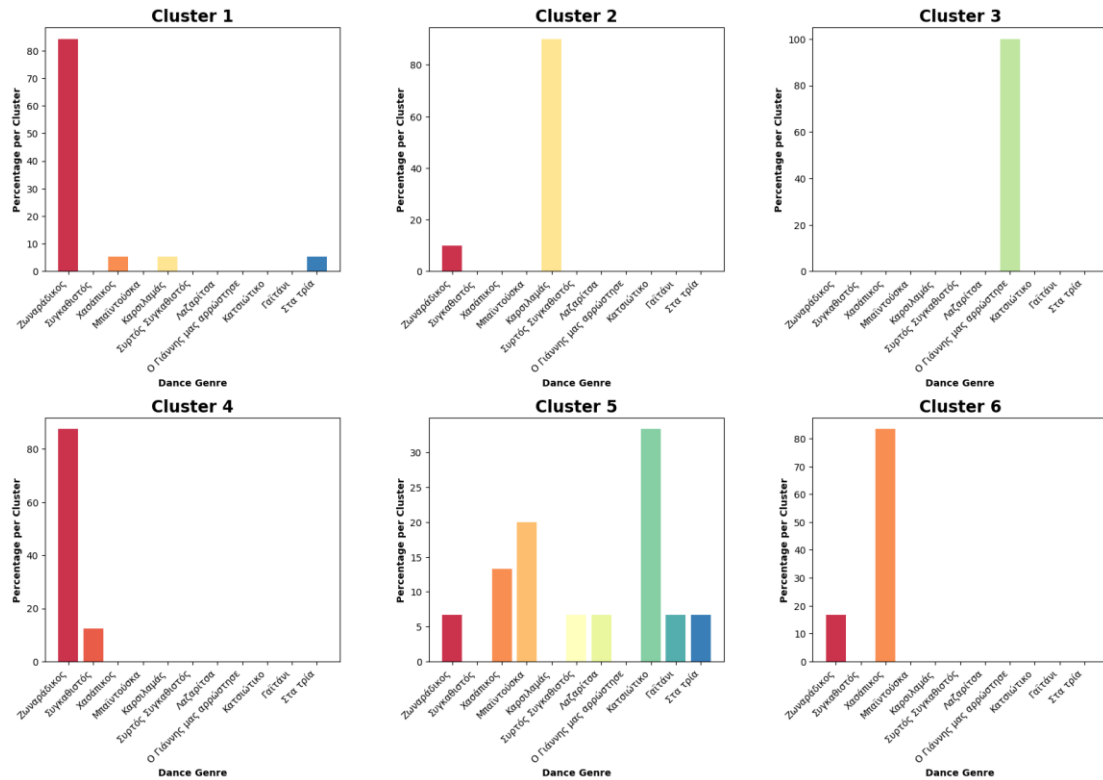
Εδώ, ακόμη περισσότερο από τα απλά μοντέλα που αναλύθηκαν παραπάνω είναι αναγκαία η χρήση της τεχνικής PCA, μειώνοντας τις διαστάσεις των διανυσμάτων χαρακτηριστικών. Και για αυτά τα μοντέλα κατασκευάστηκαν διαγράμματα *cumulative explained variance*, όπου παρότι τον αρκετά μεγαλύτερο αριθμό εξαγόμενων χαρακτηριστικών, δεν παρατηρείται αυξημένος κατά αναλογία και ο αριθμός των απαραίτητων κύριων συνιστωσών, ώστε αναπαρασταθεί μεγάλο ποσοστό της αρχικής πληροφορίας. Πιο συγκεκριμένα, φαίνεται ότι για πάνω από 60 κύριες συνιστώσες το ποσοστό της διατηρούμενης πληροφορίας αγγίζει το 90%, που είναι ένα αρκετά καλό νούμερο. Έτσι, για το συγκεκριμένο μοντέλο επιλέχθηκε  $num\_components = 63$ .



Σχήμα 5.8: Διάγραμμα cumulative explained variance για το VGG16.

Μετά τη μείωση των διαστάσεων των features, οι κύριες συνιστώσες χαρακτηριστικών που έχουν προκύψει, χρησιμοποιούνται σαν είσοδος για τον αλγόριθμο συσταδοποίησης k-means. Πρόκειται για την πρώτη απόπειρα να δοκιμαστούν τα δεδομένα σε προ-εκπαιδευμένο μοντέλο, οπότε αναμένουμε και μια βελτίωση των αποτελεσμάτων.

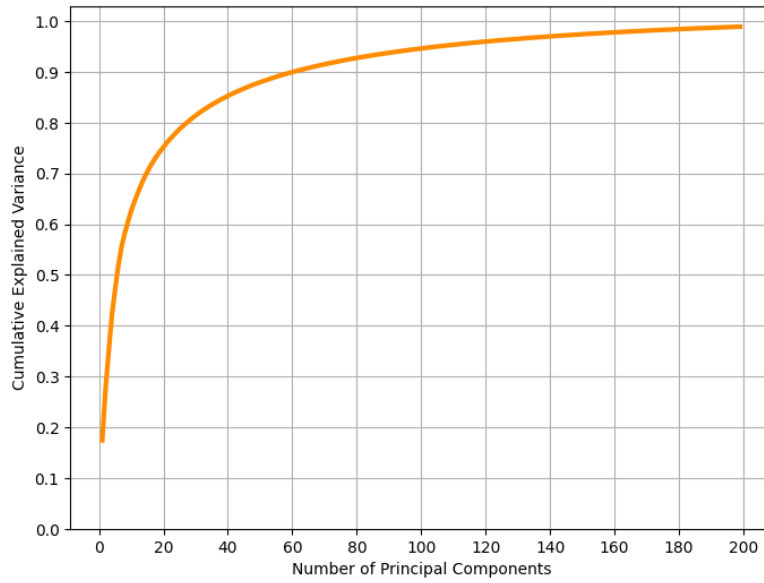
Σαν γενική παρατήρηση, έχοντας συνολική εικόνα και για τις 21 συστάδες, με τα ακριβή στοιχεία των κινησιογραμμάτων που περιλαμβάνει η καθεμία, θα ήθελα να επισημάνω ότι σε σχέση με τα τρία μοντέλα που εξετάστηκαν παραπάνω, τα clusters είναι πλέον ομοιόμορφα ως προς το πλήθος των παρτιτούρων, παρότι, ακόμη, όπως φαίνεται και στο διάγραμμα δεν υπάρχει σαφής διαχωρισμός μεταξύ των κατηγοριών του χορού. Ακόμη, θεωρώ κρίσιμο να αναφέρω ότι στις περισσότερες ομάδες, όσα είδη και αν περιλαμβάνουν, συνήθως κάποιο υπερτερεί, κάτι το οποίο θεωρώ θετικό και δε συνέβαινε στα μη προ-εκπαιδευμένα μοντέλα. Πιο αναλυτικά, στο 1<sup>ο</sup> και 4<sup>ο</sup> cluster, όπου κυρίαρχο είδος χορού είναι ο *Ζωναράδικος*, περιλαμβάνονται όλες οι παραλλαγές κάποιων χορών που ανήκουν στο είδος αυτό, όπως για παράδειγμα οι 7 παραλλαγές της *Γκίκνας* και οι 5 παραλλαγές του *Ζαβρατένου*, που δείχνει ότι υπάρχει σχετική within-cluster cohesion. Η 5<sup>η</sup> ομάδα αποτελεί ένα «πταίσμα», δεδομένης της ύπαρξης αρκετών ειδών χορού, με το *Κατσιώτικο* να υπερτερεί, αλλά όχι σε μεγάλο βαθμό. Το 6<sup>ο</sup> cluster είναι αφιερωμένο στον *Χασάπικο*, με ποσοστό μεγαλύτερο του 80%, όπου περιλαμβάνονται όλες οι εκδοχές του χορού «*Χασαπιά*». Τέλος, ιδιαίτερα ενθαρρυντικό είναι το αποτέλεσμα της 3<sup>ης</sup> συστάδας, με την κατηγορία *Ο Γιάννης μας αρρώστησε* να αγγίζει το 100% και μάλιστα, εδώ να περιλαμβάνονται και οι 7 παραλλαγές του. Παρατίθεται, λοιπόν, το διάγραμμα των ποσοστιαίων εμφανίσεων των ειδών χορού για 6 από τα clusters που δημιουργούνται.



Σχήμα 5.9: Αποτελέσματα clustering για το VGG16.

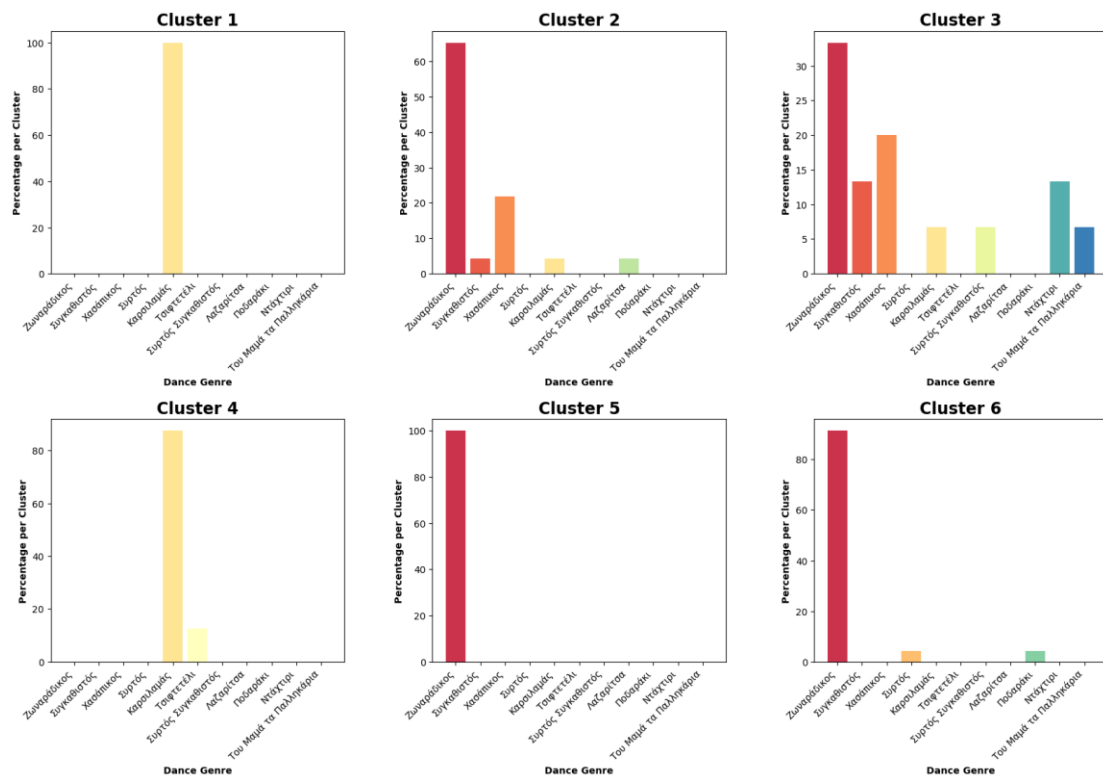
### 5.3.2 VGG19

Το αμέσως επόμενο μοντέλο που μελετάται ανήκει στην ίδια οικογένεια δικτύων, οπότε εμφανίζει και παρόμοια χαρακτηριστικά. Η βασική τους διαφορά έγκειται στο γεγονός ότι αποτελείται από 19 κύρια επίπεδα: 16 convolutional (ανάμεσα στα οποία παρεμβάλλονται και max pooling layers) και 3 fully connected. Ακολουθήθηκε ίδια μεθοδολογία με αυτή που περιγράφηκε παραπάνω. Τελικά, προέκυψαν 271 διανύσματα χαρακτηριστικών μεγέθους  $1 \times 4096$ . Παρακάτω φαίνεται και το διάγραμμα που αναπαριστά την *cumulative explained variance* για το VGG19, που βοήθησε στην επιλογή αριθμού κυρίων συνιστωσών. Η αντιστοιχία αριθμού principal components με το ποσοστό της αρχικής πληροφορίας που διατηρείται προσομοιάζει αυτή του VGG16. Ωστόσο, μετά από αρκετούς πειραματισμούς και δοκιμές, κυρίως σε σχέση με τα τελικά αποτελέσματα που προέκυψαν μετά την εφαρμογή του αλγορίθμου συσταδοποίησης, καταλήξαμε στο ότι θα χρειαστούμε  $num\_components = 154$ , για καλύτερα αποτελέσματα.



Σχήμα 5.10: Διάγραμμα cumulative explained variance για το VGG19.

Στη συνέχεια, σειρά έχει η συσταδοποίηση των δεδομένων, οπότε εφαρμόζεται ο αλγόριθμος k-means και τα αποτελέσματα 6 εκ των συστάδων φαίνονται στο παρακάτω σχήμα.



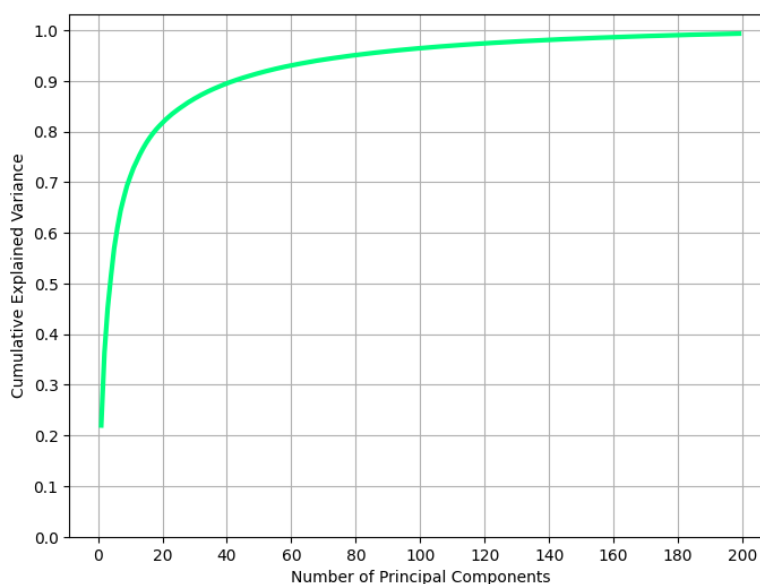
Σχήμα 5.11: Αποτελέσματα clustering για το VGG19.

Εδώ βλέπουμε ότι τα συγκεκριμένα clusters είναι καλύτερα διαχωρισμένα σε σχέση με τα αποτελέσματα του VGG16 και, πράγματι, δεν είναι τυχαίο, ακολουθείται περίπου το

ίδιο μοτίβο και για τα υπόλοιπα 15 clusters. Βάσει του παραπάνω σχήματος, εξαίρεση, θα μπορούσαμε να πούμε ότι, αποτελεί η 3<sup>η</sup> ομάδα, όπου περιλαμβάνονται και πάλι αρκετά είδη χορού, χωρίς να ξεχωρίζει κάποιο. Ωστόσο, στο 1<sup>ο</sup> και 5<sup>ο</sup> bar plot, παρατηρούμε ότι ο *Καρσιλαμάς* και ο *Ζωναράδικος* αντίστοιχα φτάνουν το 100%, οπότε, οι εν λόγω συστάδες περιλαμβάνουν αποκλειστικά παραλλαγές χορών που ανήκουν στα συγκεκριμένα είδη, κάτι το οποίο είναι ιδιαίτερος θετικό. Όσον αφορά το 2<sup>ο</sup> και 6<sup>ο</sup> cluster, είναι προφανές ότι υπερτερεί και πάλι ο *Ζωναράδικος*, παρά το αυξημένο πλήθος κινησιογραμμάτων που περιλαμβάνει το καθένα (περισσότερα από 15).

### 5.3.3 ResNet50

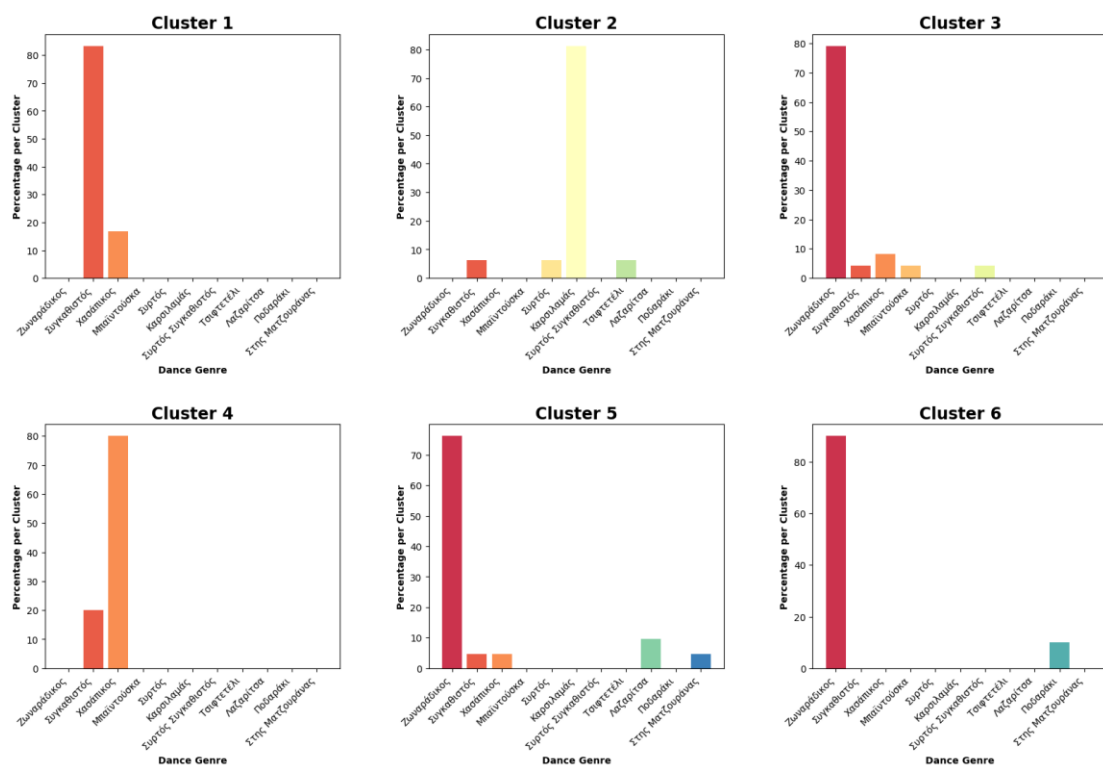
Προχωρώντας στο επόμενο μοντέλο που εξετάστηκε στα πλαίσια της διπλωματικής αυτής εργασίας, πρόκειται για ένα νευρωνικό δίκτυο με αρχιτεκτονική πολύ βαθύτερη από τα δύο παραπάνω. Χρησιμοποιεί τα λεγόμενα residual blocks, τα οποία έχουν περιγραφεί αναλυτικά στο τρίτο κεφάλαιο και αποτελεί μια πολύ καλή επιλογή ως εξαγωγέας χαρακτηριστικών στην περίπτωση μας, καθώς, εν γένει, εμφανίζει υψηλή απόδοση ακόμη και για μικρά σύνολα δεδομένων. Αποτελείται συνολικά από 50 κύρια επίπεδα: 49 convolutional και 1 fully connected, για την πρόβλεψη των κλάσεων. Ανάμεσα στα συνελκτικά στρώματα συχνά παρεμβάλλονται batch normalization layers, ενώ σε όλο το δίκτυο υπάρχουν μόνο δύο συγκεντρωτικά επίπεδα, ένα στην αρχή και ένα στο τέλος. Σαν αποτέλεσμα λαμβάνουμε 271 διανύσματα χαρακτηριστικών μεγέθους  $1 \times 2048$ .



Σχήμα 5.12: Διάγραμμα cumulative explained variance για το ResNet50.

Τα διανύσματα αυτά υφίστανται μείωση διαστατικότητας, μέσω της μεθόδου της ανάλυσης κυρίων συνιστωσών, όπου βοηθητικό ρόλο έχει η *cumulative explained variance*, το διάγραμμα της οποίας απεικονίζεται παραπάνω. Εδώ, βλέπουμε ότι ήδη με αριθμό components ίσο με 40 επιτυγχάνεται ποσοστό 90% της αρχικής πληροφορίας, ενώ και ο αριθμός 1.0 προσεγγίζεται συντομότερα από τα VGG16 και VGG19. Έτσι, σε αυτήν την περίπτωση διαλέξαμε  $num\_components = 62$ .

Επόμενο βήμα είναι το πέρασμα των δεδομένων από τον αλγόριθμο k-means, ο οποίος θα τα ομαδοποιήσει σε 21 συνολικά συστάδες. Παρακάτω, δίδεται διάγραμμα που περιλαμβάνει 6 από τις συστάδες, με τις κατηγορίες χορών που περιλαμβάνει η καθεμία.



Σχήμα 5.13: Αποτελέσματα clustering για το ResNet50.

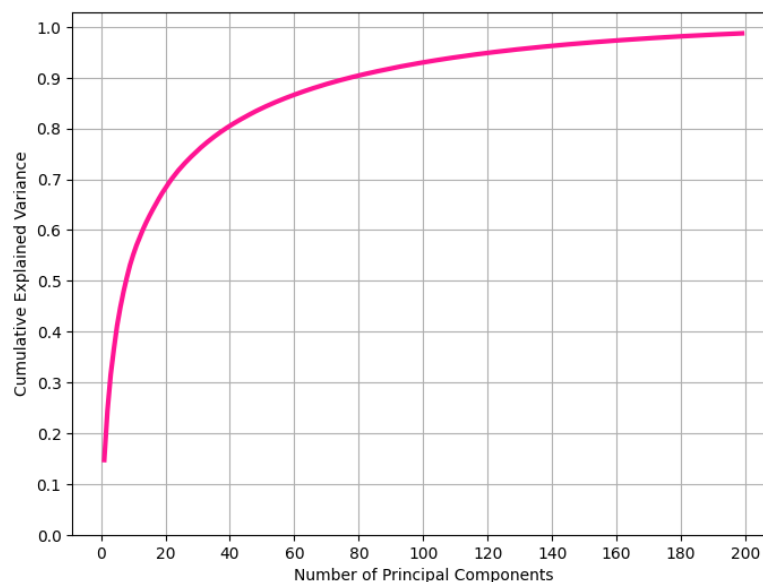
Σύμφωνα με τα παραπάνω διαγράμματα, αλλά και τις μετρικές αξιολόγησης, το δίκτυο αυτό μας δίνει τα μέχρι στιγμής καλύτερα αποτελέσματα, όσον αφορά το κομμάτι της συσταδοποίησης. Πράγματι, όπως βλέπουμε πιο πάνω, όλες οι συστάδες, παρότι δεν περιλαμβάνουν ένα μόνο είδος χορού, εμφανίζουν μεγαλύτερη συγκέντρωση κινήσιωγραμμάτων σε κάποια συγκεκριμένη κατηγορία, το οποίο θεωρείται σημαντική βελτίωση. Σε 3 από τα clusters υπερτερεί ο Ζωναράδικος και πιο συγκεκριμένα, στο 5<sup>ο</sup>, έχουμε όλες τις παραλλαγές από 4 διαφορετικούς χορούς που ανήκουν σε αυτό το είδος, κάτι το οποίο δεν έχουμε ξανασυναντήσει νωρίτερα, ενώ στο 6<sup>ο</sup> έχουμε και τις 7 εκδοχές του Ζωναράδικου Τσέστου. Στην 4<sup>η</sup> συστάδα περιλαμβάνονται όλες οι παραλλαγές της



Χασαπιάς, όπως είχε καταφέρει να ομαδοποιήσει και το VGG16 και, τέλος, στη 2<sup>η</sup> συστάδα που είναι αφιερωμένη στον *Καρσιλαμά*, θα ήθελα να αναφέρω ότι τα άλλα τρία είδη που φαίνονται έχουν ένα μόνο χορό στο συγκεκριμένο cluster, ενώ όλα τα υπόλοιπα κινησιογράμματα ανήκουν στο επικρατές είδος. Επομένως, συνολικά προκύπτει μια καλή εικόνα όσον αφορά τη συσταδοποίηση που έγινε στα δεδομένα, τα οποία εξήγαγε το συγκεκριμένο μοντέλο.

### 5.3.4 InceptionV3

Αμέσως επόμενο στη σειρά είναι το μοντέλο InceptionV3, που εισήγαγε τη φιλοσοφία των πλατιών και όχι πια βαθιών δικτύων για βελτίωση των αποτελεσμάτων. Η οικογένεια αυτών των νευρωνικών δικτύων χρησιμοποιεί τα inception modules. Η αρχιτεκτονική που προτείνει είναι, πράγματι, πιο ελαφριά σε σχέση με άλλα δίκτυα, πράγμα το οποίο του δίνει πλεονέκτημα σχετικά με την υπολογιστική ισχύ, καθώς και τις απαιτήσεις σε μνήμη. Ακόμη, θεωρείται ένα εξαιρετικά αποτελεσματικό δίκτυο που έχει επιτύχει πολύ υψηλές επιδόσεις σε ποικιλία προβλημάτων του πεδίου που εξετάζουμε. Όσον αφορά τον αριθμό των επιπέδων που το απαρτίζουν, ισούται συνολικά με 48 επίπεδα: 17 συνελκτικά, 3 πλήρως συνδεδεμένα και τα υπόλοιπα είναι συγκεντρωτικά και κανονικοποίησης δέσμης. Τελικά, από το μοντέλο αυτό προκύπτουν διανύσματα μεγέθους  $1 \times 2048$  για κάθε μία από τις 271 εικόνες του συνόλου δεδομένων.

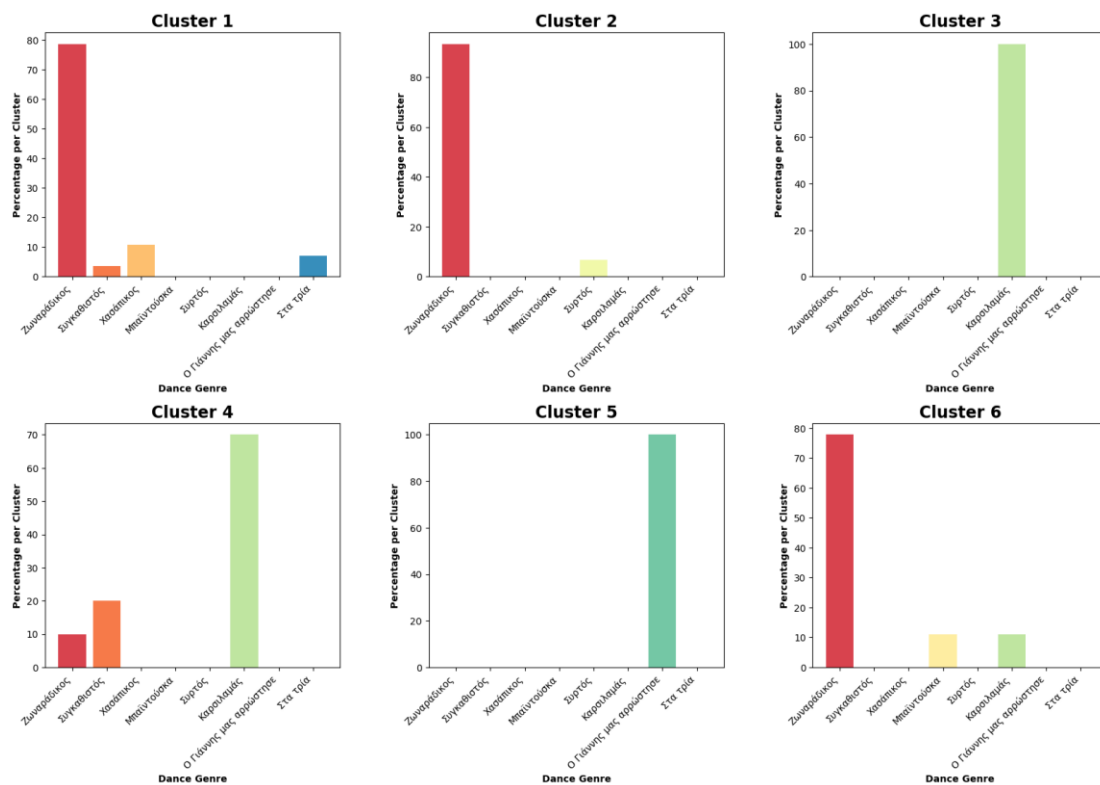


Σχήμα 5.14: Διάγραμμα cumulative explained variance για το InceptionV3.

Παραπάνω φαίνεται και το διάγραμμα της *cumulative explained variance*, όπου παρατηρούμε ότι απαιτείται ο διπλάσιος αριθμός κυρίων συνιστωσών, σε σχέση με το

ResNet50, ώστε να φτάσουμε τουλάχιστον στο 90% της αρχικά παρεχόμενης πληροφορίας. Παρόλα αυτά, δοκιμές που πραγματοποιήθηκαν με αριθμούς που κυμαίνονταν ανάμεσα στο 80 και το 120, παρήγαγαν χειρότερα αποτελέσματα σε σχέση με μικρότερες τιμές για τον αριθμό των principal components. Επομένως, συμβουλευόμενοι πάντα τα τελικά αποτελέσματα της συσταδοποίησης, καταλήξαμε στο  $num\_components = 45$ .

Ακολουθεί ο εξαιρετικά δημοφιλής αλγόριθμος k-means, με στόχο να ομαδοποιήσει τα δεδομένα που του παρέχονται σε 21 διαφορετικές συστάδες. Έτσι, παρακάτω, σε μια προσπάθεια οπτικοποίησης των αποτελεσμάτων, δίνονται 6 bar plots, καθένα από τα οποία αντιπροσωπεύει ένα cluster και αντιστοίχως φαίνονται οι κατηγορίες που περιέχει.



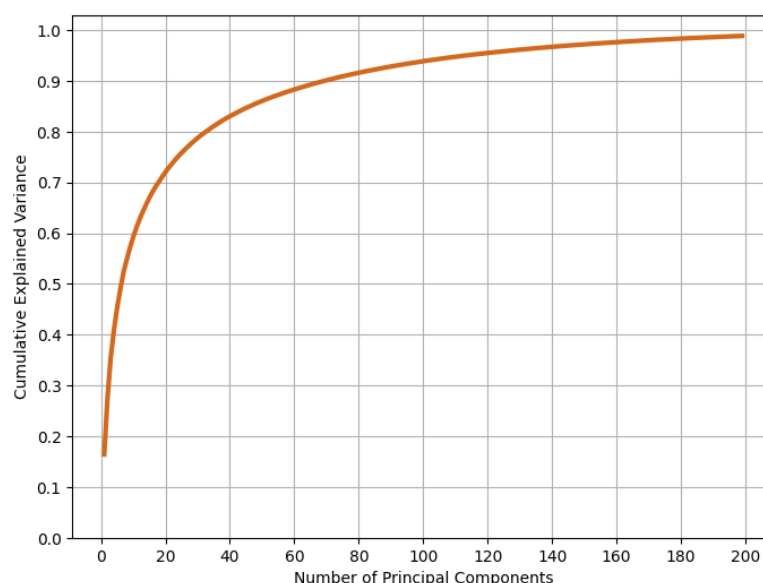
Σχήμα 5.15: Αποτελέσματα clustering για το InceptionV3.

Πρόκειται, επίσης, για ένα μοντέλο που προσέφερε υψηλά αποτελέσματα, (βάσει πάντοτε των τιμών που προέκυψαν για το συγκεκριμένο πρόβλημα που εξετάζεται στην εργασία) και, όπως παρατηρούμε από τα διαγράμματα, δίνει, επίσης, καλά διαχωρισμένες κλάσεις, εν συγκρίσει με τα υπόλοιπα μοντέλα. Δεν βλέπουμε να ξεπερνούν τα 4 είδη ανά συστάδα, το οποίο είναι αρκετά αντιπροσωπευτικό, καθώς ισχύει και για τα υπόλοιπα clusters που δεν απεικονίζονται, με εξαίρεση 2, τα οποία περιέχουν 6 και 7 κατηγορίες αντίστοιχα. Φυσικά, εάν αυτοί οι αριθμοί συγκριθούν με άλλα μοντέλα, είναι σαφώς βελτιωμένοι, καθώς παρατηρούνταν ακόμη και μεμονωμένες περιπτώσεις συστάδων που

περιείχαν 9 ή 10 κατηγορίες. Περνώντας σε πιο στοχευμένες παρατηρήσεις, θα ήθελα, αρχικά, να αναφέρω ότι στο 1<sup>ο</sup> και 6<sup>ο</sup> διάγραμμα, όπου υπερτερεί ο *Ζωναράδικος* και πάλι, είναι συγκεντρωμένα κινησιογράμματα από παραλλαγές συγκεκριμένων χορών, πράγμα το οποίο μας δείχνει ότι αναγνωρίστηκε η ομοιότητα μεταξύ ορισμένων χορών. Επίσης, στην 4<sup>η</sup> συστάδα περιλαμβάνονται όλες οι διαφορετικές εκδοχές του χορού «*Το Λενάκι πλάθει πίτα*» που ανήκει στο είδος του *Καρσιλαμά*. Τέλος, η 5<sup>η</sup> συστάδα είναι πλήρως καλυμμένη από τον χορό «*Ο Γιάννης μας αρρώστησε*» και όλες τις παραλλαγές του, όπως συμβαίνει και στο δίκτυο VGG16.

### 5.3.5 Xception

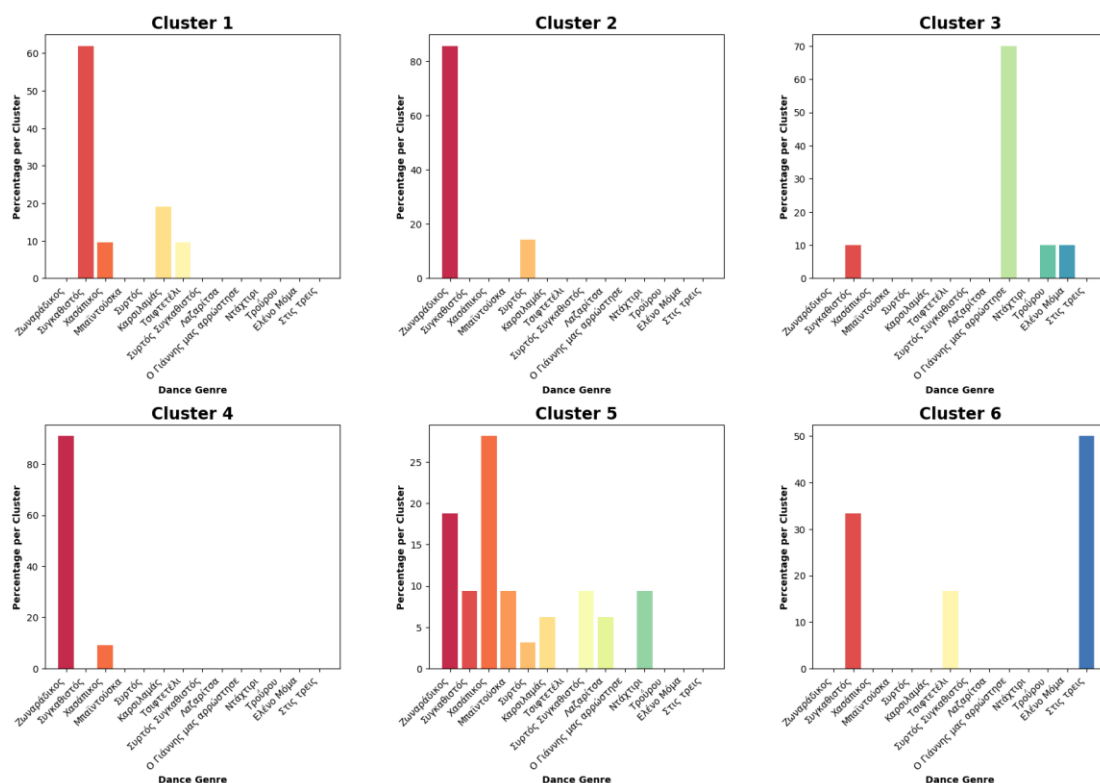
Το τελευταίο από τα μοντέλα που μελετήθηκαν για την υλοποίηση της εργασίας είναι το Xception, του οποίου η ιδιαιτερότητα έγκειται στη χρήση των κατά βάθος διαχωρίσιμων συνελίξεων (Depthwise Separable Convolutions). Έχει γνωρίσει μεγάλη επιτυχία, χρησιμοποιούμενο σε ευρύ φάσμα εφαρμογών, λόγω της εξαιρετικής του απόδοσης σε συνδυασμό με το μειωμένο αριθμό παραμέτρων, που απαιτεί και μικρότερη υπολογιστική ισχύ. Συνολικά, αποτελείται από 36 συνελκτικά επίπεδα. Ως εξαγωγέας χαρακτηριστικών παράγει 271 διανύσματα μεγέθους  $1 \times 2048$ , τα οποία στη συνέχεια μεταβιβάζονται στο στάδιο μείωσης της διαστατικότητας PCA. Το διάγραμμα της *cumulative explained variance* παρουσιάζεται παρακάτω, όπου ο αριθμός των principal components είναι και πάλι αυξημένος σε σχέση με του ResNet50, όμως ακολουθεί παρόμοια πορεία με του InceptionV3. Και σε αυτήν την περίπτωση, παρότι ο αριθμός των κυρίων συνιστωσών που



Σχήμα 5.16: Διάγραμμα *cumulative explained variance* για το Xception.

απεικονίζει πάνω από το 90% της πληροφορίας είναι τουλάχιστον 70, βάσει των πειραμάτων που έγιναν αποδείχθηκε ότι τα καλύτερα αποτελέσματα τα λάβαμε για  $num\_components = 36$ .

Στο σημείο αυτό, έρχεται η σειρά της συσταδοποίησης, όπου με την εφαρμογή του αλγορίθμου k-means επιτυγχάνεται και ο τελικός σκοπός της εργασίας. Παρακάτω παρατίθενται τα αποτελέσματα για 6 από τις συστάδες που δημιουργήθηκαν βάσει των χαρακτηριστικών που εξήγαγε το μοντέλο Xception.



Σχήμα 5.17: Αποτελέσματα clustering για το Xception.

Το μοντέλο αυτό, σε αντίθεση με τα 2 προηγούμενα, δεν πετυχαίνει τόσο υψηλή απόδοση, το οποίο φαίνεται και από το δείγμα των συστάδων που παρουσιάζονται παραπάνω. Αρχικά, στο 5<sup>ο</sup> cluster, βλέπουμε ότι υπάρχουν 9 διαφορετικά είδη χορού, χωρίς κάποιο από αυτά να είναι κυρίαρχο, το οποίο δεν είναι ιδιαίτερος θεμιτό για το πρόβλημα που εξετάζουμε. Παρόλα αυτά, τα υπόλοιπα clusters, φαίνονται καλύτερα διαχωρισμένα. Στην 1<sup>η</sup> συστάδα, περιλαμβάνονται αρκετές παραλλαγές του *Καρσιλαμάς*, καθώς και όλες οι εκδοχές του χορού «*Συγκα' στό του γάμου*» που ανήκουν στο γένος *Συγκαθιστός*, ο οποίος έχει και τη μεγαλύτερη συχνότητα εμφάνισης. Στο 2<sup>ο</sup> και 4<sup>ο</sup> cluster, έχουμε ξανά «νίκη» του *Ζωναράδικου*, με παραλλαγές των χορών «*Σουλτάνα Σουφλιουτούδα*» και «*Ζαβρατένο*» να συγκεντρώνονται όλες μαζί στο κάθε cluster αντίστοιχα (ο ένας χορός στο ένα και ο άλλος στο δεύτερο). Στην 3<sup>η</sup> συστάδα, βλέπουμε

ότι υπάρχουν όλες οι παραλλαγές του χορού «*Ο Γιάννης μας αρρώστησε*», ωστόσο, με βάση τα χαρακτηριστικά που εξήγαγε το συγκεκριμένο δίκτυο ο αλγόριθμος δεν κατάφερε να τα αντιστοιχίσει σε ξεχωριστή συστάδα, οπότε συνυπάρχουν με 3 ακόμη είδη. Τέλος, στην 6<sup>η</sup> ομάδα, υπάρχουν όλες οι παραλλαγές του χορού «*Στις τρεις*», μαζί με 2 κινησιογράμματα που ανήκουν στην κατηγορία του *Συγκαθιστού* και ένα, ακόμη, στο είδος *Τσιφτετέλι*.

Συνοψίζοντας, θα ήθελα να αναφερθώ σε ορισμένες συνολικές παρατηρήσεις που αφορούν τα παραπάνω μοντέλα. Αρχικά, φαίνεται πως ο Ζωναράδικος υπερτερεί σε όλα τα διαγράμματα, κάτι που σημαίνει ότι για κάθε περίπτωση που μελετάται στην παρούσα εργασία, η κατηγορία αυτή εμφανίζεται με μεγάλη συγκέντρωση σε αρκετές συστάδες. Αυτό εξηγείται, ενδεχομένως, λόγω του γεγονότος ότι το συγκεκριμένο είδος χορού περιλαμβάνει σχεδόν τα μισά κινησιογράμματα του συνόλου δεδομένων (114 από τα 271 συνολικά), οπότε για τον εκάστοτε αλγόριθμο αποτελεί πρόκληση η ομαδοποίησή των χορών στις σωστές συστάδες. Αυτό έχει σαν αποτέλεσμα, η *majority class* – Ζωναράδικος να διαχωρίζεται σε επιμέρους clusters, στην προσπάθεια του k-means να ελαχιστοποιήσει τη συνολική διακύμανση και να παράγει αποτελέσματα πιο κοντά στον στόχο της συσταδοποίησης. Επιπλέον, παρατήρησα στα προ-εκπαιδευμένα μοντέλα ότι σε ορισμένους χορούς, όπως «*Ο Γιάννης μας αρρώστησε*», «*Το Λενάκι πλάθει πίτα*», «*Ζωναράδικος Τσέστος*», «*Γκίκνα*», «*Ζαβρατένο*» οι παραλλαγές τους ομαδοποιούνται στην ίδια συστάδα, με μεγάλη συχνότητα, δηλαδή σε όλα (σχεδόν) τα μοντέλα, κάτι το οποίο είναι θετικό και σημαίνει ότι οι συγκεκριμένοι χοροί έχουν ορισμένα χαρακτηριστικά που ξεχωρίζουν και έτσι, γίνονται πιο εύκολα διαχωρίσιμοι από άλλους. Τέλος, αναφερόμενη και πάλι στα συνελκτικά δίκτυα που αξιοποιήθηκε η μεταφορά μάθησης, και ακόμη περισσότερο στα δύο μοντέλα που έφεραν τα καλύτερα αποτελέσματα (ResNet50, InceptionV3), θα ήθελα να τονίσω ότι παρότι οι συστάδες δεν ήταν καλά διαχωρισμένες, υπό την έννοια ότι μπορεί να περιλάμβαναν αρκετές κατηγορίες χορού η καθεμία, σχεδόν πάντα υπήρχε επικρατέστερη κατηγορία με μεγάλη συγκέντρωση κινησιογραμμάτων, που σημαίνει ότι υπήρχε, εν γένει, συνεκτικότητα εντός συστάδας (within-cluster cohesion), και, ίσως, όχι ιδιαίτερος αυξημένος διαχωρισμός μεταξύ των συστάδων (between-cluster separation).

## 5.4 Συγκεντρωτικά αποτελέσματα

Στο σημείο αυτό της εργασίας γίνεται παρουσίαση των τελικών αποτελεσμάτων, όπως αυτά προκύπτουν αξιολογώντας τα ως προς την εγκυρότητα της συσταδοποίησης με

διαφορετικές μετρικές. Φυσικά, παραθέτουμε τα βέλτιστα μόνο αποτελέσματα, τα οποία εξήχθησαν βάσει συγκεκριμένων υπερπαραμέτρων που χρησιμοποιήθηκαν σε κάθε πείραμα. Το πρώτο συνελκτικό μοντέλο που χρησιμοποιήθηκε αναφέρεται ως CNN1, το δεύτερο ως CNN3 (λόγω του ότι περιλαμβάνει 3 συνελκτικά επίπεδα) και το τρίτο ως CNN4 (λόγω του ότι περιλαμβάνει 4 συνελκτικά επίπεδα).

Να υπενθυμίσουμε, επίσης, ότι οι τρεις πρώτες μετρικές (rand, nmi, purity score) χρησιμοποιούν τόσο τα ground truth, όσο και τα predicted labels και παίρνουν τιμές από 0 έως 1, με το 0 να αντιπροσωπεύει λανθασμένη συσταδοποίηση και το 1 πλήρως επιτυχημένη. Αντιθέτως, η τέταρτη μετρική (davies-bouldin score), χρησιμοποιεί μόνο τα ίδια τα δεδομένα και τα predicted labels, λαμβάνοντας τιμές από το 0, χωρίς να υπάρχει κάποιο ανώτατο όριο, και όσο μικρότερη είναι η τιμή (κοντά στο 0), τόσο καλύτερη και η επίδοση του αλγορίθμου. Για αυτό το λόγο, διαχωρίζουμε οριζόντια τον πίνακα με διπλή γραμμή, ώστε να μην υπάρξει σύγχυση σχετικά με την ερμηνεία των τιμών των διαφορετικών μεταβλητών, μιας και όσο τα αποτελέσματα βελτιώνονται, αυτές «κινούνται» αντίθετα (οι rand, nmi, purity score → αυξάνονται, ενώ η davies-bouldin score → μειώνεται). Στον παρακάτω πίνακα τα καλύτερα αποτελέσματα για κάθε μετρική σημειώνονται με έντονο χρώμα.

Πίνακας 5.1: Συγκεντρωτική αξιολόγηση πειραμάτων.

	<i>CNN Μοντέλο</i>	<i>rand_score</i>	<i>NMI score</i>	<i>purity score</i>	<i>DB score</i>
<i>from scratch</i>	CNN1	0.6993	0.1937	0.4556	2.31
	CNN3	0.7389	0.2702	0.4889	2.15
	CNN4	0.7541	0.3053	0.5185	2.11
<i>Transfer Learning</i>	VGG16	0.7706	0.3826	0.5852	1.97
	VGG19	0.7724	0.4002	0.5889	2.06
	ResNet50	<b>0.7756</b>	0.4045	<b>0.6037</b>	<b>1.90</b>
	InceptionV3	0.7742	<b>0.4059</b>	0.5778	2.03
	Xception	0.7629	0.3903	0.5741	1.93

Αρχικά, παρατηρούμε ότι το μοντέλο που έφερε τα καλύτερα αποτελέσματα, σύμφωνα με τις 3 από τις 4 μετρικές, είναι το ResNet50, με το *normalized mutual information (NMI)* score του να απέχει μόλις λίγα χιλιοστά από την απόδοση του InceptionV3, η οποία είναι η υψηλότερη σύμφωνα με αυτή τη μετρική. Επίσης, είναι οφθαλμοφανές, όπως και αναμενόταν, ότι τα πιο απλά Συνελικτικά Δίκτυα πετυχαίνουν χαμηλότερο σκορ, με αυτό να βελτιώνεται όσο αυξάνονται και τα συνελικτικά στρώματα. Φυσικά, τα προ-εκπαιδευμένα μοντέλα έχουν καλύτερη επίδοση, με το ResNet50 και InceptionV3 να ξεχωρίζουν ελαφρώς, χωρίς όμως να παρατηρείται μεγάλη απόκλιση στα αποτελέσματα που παράγονται και από τα υπόλοιπα τρία μοντέλα.

Ωστόσο, φαίνεται ότι οι 4 αυτές μέθοδοι αξιολόγησης δίνουν αποτελέσματα που έχουν σχετικά μεγάλη απόκλιση μεταξύ τους. Όπως έχει ήδη εξηγηθεί, καθεμία από αυτές υπολογίζεται με διαφορετικό τρόπο και λαμβάνει υπόψη της διαφορετικά στοιχεία που σχετίζονται με τη συσταδοποίηση γι' αυτό και η επιλογή εξαρτάται από το εκάστοτε πρόβλημα και το τι θέλουμε να αξιολογήσουμε κάθε φορά. Να υπενθυμίσουμε πως η αξιολόγηση της επίδοσης ενός αλγορίθμου συσταδοποίησης δεν είναι κάτι απλό και κυρίως «αντικειμενικό», όπως συμβαίνει στην περίπτωση της Επιβλεπόμενης Μάθησης. Πράγματι, οι μετρικές που χρησιμοποιούνται για τέτοιου είδους προβλήματα δεν συγκρίνουν απλώς τα *ground truth* με τα *predicted labels*, αλλά ακολουθώντας διαφορετικές μαθηματικές μεθόδους η καθεμία, δίνουν ένα αποτέλεσμα ενδεικτικό της ποιότητας του clustering (*clustering quality*). Αυτό συμβαίνει, άλλοτε χρησιμοποιώντας τα *true labels*, εφόσον αυτά υπάρχουν, και άλλοτε χωρίς τις ετικέτες, αξιολογώντας απλώς το εάν τα δείγματα που έχουν ταξινομηθεί στην ίδια κλάση μοιάζουν περισσότερο μεταξύ τους, παρά με εκείνα που βρίσκονται σε διαφορετικές.

Ωστόσο, θα ήθελα να επισημάνω πως, παρότι οι 4 μετρικές δεν κυμαίνονται στα ίδια ποσοστά επιτυχίας, κάτι το οποίο περιμέναμε, όπως εξηγήσαμε παραπάνω, βλέπουμε ότι συμφωνούν, σε γενικές γραμμές, και ότι κινούνται σε κοινό άξονα ως προς την αξιολόγηση της συσταδοποίησης, πλην ορισμένων διαφορών, κάτι το οποίο μας βοηθάει στην εξαγωγή συμπερασμάτων.

Πιο αναλυτικά, παρατηρούμε ότι η μετρική *rand\_score* πετυχαίνει συνολικά τα υψηλότερα ποσοστά επιτυχίας, συγκρινόμενη με το *nmi* και *purity score*. Ωστόσο, θα ήθελα να αναφέρω ότι στην περίπτωση των μη προ-εκπαιδευμένων μοντέλων, κυρίως, οι τιμές αυτές δε φαίνεται να συμφωνούν με τον διαχωρισμό σε *clusters* των δεδομένων, εφόσον υπάρχει μεγάλη ανομοιομορφία τόσο ως προς τα είδη χωρών που περιέχονται σε κάθε συστάδα, όσο και ως προς το πλήθος των κινησιογραμμάτων που περιλαμβάνει

καθεμία (ορισμένα clusters περιλαμβάνουν ένα ή δύο κινησιογράμματα και άλλα περισσότερα από 50!). Όσον αφορά τα υπόλοιπα μοντέλα, θεωρώ ότι τα αποτελέσματα που δίνει αν και ελαφρώς αυξημένα, πλησιάζουν τις πραγματικές επιδόσεις του αλγορίθμου, έχοντας εξετάσει ενδελεχώς τα κινησιογράμματα που περιλαμβάνονται σε κάθε συστάδα.

Επόμενη μετρική που χρησιμοποιείται είναι το nmi score, τα αποτελέσματα του οποίου εκ πρώτης όψης δεν είναι ιδιαίτερος ικανοποιητικά. Ας μην ξεχνάμε, ωστόσο, ότι πρόκειται για μια μη γραμμική μετρική αξιολόγησης, η οποία είναι κάπως πιο «αυστηρή», όπως έχει αναφερθεί στην προηγούμενη υποενότητα και σύμφωνα με το [Σχήμα 5.1](#) η τιμή 0.4 που προκύπτει για τα καλύτερα μοντέλα μας, αντιστοιχεί σε ποσοστό μεγαλύτερο του 80%, το οποίο αντιπροσωπεύει το ποσοστό των δειγμάτων που έχουν ομαδοποιηθεί σωστά. Επιπλέον, να αναφέρουμε ότι, εν γένει, το NMI τείνει να αυξάνεται όσο αυξάνεται η αναλογία  $N/K$ , όπου  $N$  είναι ο αριθμός των δειγμάτων και  $K$  ο αριθμός των συστάδων [122]. Στην περίπτωση μας, η αναλογία αυτή δεν είναι ιδιαίτερος μεγάλη, οπότε το NMI θεωρείται μια αξιόπιστη μετρική για το πρόβλημα που μελετάμε.

Αμέσως, μετά επιλέχθηκε το purity score, λόγω της απλότητας και της ευκολίας του ως προς την ερμηνεία. Τα ποσοστά που παράγει για τα προ-εκπαιδευμένα μοντέλα, τα θεωρώ κοντά στην πραγματικότητα, ίσως, ελαφρώς πιο χαμηλά, σε αντίθεση με τα ποσοστά των *CNN1*, *CNN3* και *CNN4*, τα οποία είναι πιο υψηλά από ότι θα έπρεπε. Η αλήθεια είναι ότι η εν λόγω μετρική έχει αρκετούς περιορισμούς, ο πιο σημαντικός από τους οποίους είναι ότι δίνει σωστότερα αποτελέσματα για δεδομένα που σχηματίζουν σαφείς και καλά ορισμένες συστάδες, κάτι το οποίο δε συμβαίνει στην περίπτωση που εξετάζουμε. Ωστόσο, αποτελεί μια εν γένει διαδεδομένη μετρική αξιολόγησης για την εγκυρότητα της συσταδοποίησης, οπότε στην εν λόγω εργασία χρησιμοποιείται βοηθητικά και σε συνδυασμό με άλλες μετρικές, ώστε να εξαχθούν τα αντίστοιχα συμπεράσματα.

Τελευταία μετρική αξιολόγησης που χρησιμοποιείται στην παρούσα εργασία είναι το davies-bouldin score, του οποίου τα αποτελέσματα είναι εντελώς διαφορετικά από τις υπόλοιπες τρεις μετρικές, όσον αφορά την τιμή που παράγεται. Παρόλα αυτά, θα μπορούσαμε να πούμε ότι οι τιμές αυτές είναι λογικές, καθώς πρόκειται για αποτελέσματα που πλησιάζουν το 0, το οποίο και αντιπροσωπεύει την «τέλεια» συσταδοποίηση. Θα ήθελα, ακόμη, να παρατηρήσω ότι υπάρχουν κάποιες διαφοροποιήσεις ως προς την απόδοση των προ-εκπαιδευμένων μοντέλων σε σύγκριση με τις πρώτες τρεις μετρικές. Πιο συγκεκριμένα, φαίνεται ότι, σύμφωνα με την εν λόγω μέθοδο αξιολόγησης, τα μοντέλα *VGG16* και *Xception* παράγουν καλύτερα αποτελέσματα από τα *VGG19* και *InceptionV3*,



κάτι το οποίο την φέρνει σε αντίθεση με τις *rand*, *nmi* και *purity score*. Βέβαια, οι διαφορές αυτές, μεταξύ των αποτελεσμάτων είναι τόσο μικρές, που θα μπορούσαν να χαρακτηριστούν αμελητέες και, ίσως, με προσαρμογή των υπερπαραμέτρων, ακόμη και αυτή η μικρή απόκλιση να μην υφίσταντο.

Συνολικά, θα ήθελα να επισημάνω ότι για τα πειράματα χρησιμοποιήθηκε το συγκεκριμένο σύνολο δεδομένων, το οποίο, όπως έχουμε αναφέρει και στο 4<sup>ο</sup> κεφάλαιο, θεωρείται *imbalanced*, κάτι το οποίο, ενδεχομένως εξηγεί εν μέρει τα αποτελέσματα. Πράγματι, μερικά είδη χορών περιλαμβάνουν μόνο 2 ή 3 κινησιογράμματα από τα 271 συνολικά, ενώ η κατηγορία *Ζωναράδικος* που περιλαμβάνει τα περισσότερα, αριθμεί 114 παρτιτούρες, με τις αμέσως επόμενες σε πλήθος κατηγορίες να είναι ο *Συγκαθιστός* που περιέχει 32, ο *Καρσιλαμάς* με 29 και ο *Χασάπικος* με 28. Συνεπώς, έχουμε συγκέντρωση 204 κινησιογραμμάτων σε 4 κατηγορίες, και απομένουν 67 μόνο κινησιογράμματα, τα οποία κατανέμονται στις υπόλοιπες 17 κατηγορίες. Όπως καταλαβαίνουμε, το γεγονός αυτό επηρεάζει σημαντικά τα τελικά αποτελέσματα και, μάλιστα, με δύο τρόπους. Αρχικά, ο ίδιος ο αλγόριθμος συσταδοποίησης «δυσκολεύεται» να καταλήξει σε σωστό διαχωρισμό, λόγω αυτής της ανισορροπίας, αλλά, επίσης, όσον αφορά τις μετρικές αξιολόγησης, οι οποίες εξαρτώνται από αυτήν την ομαδοποίηση των δειγμάτων, ενδέχεται να καταλήξουν σε ανακριβείς τιμές, εφόσον ο τρόπος υπολογισμού τους, πολλές φορές μπορεί να ευνοεί τη *majority class* και αντιστοίχως να αγνοεί τις *minority classes*.

Σε όλα αυτά, θα πρέπει να προστεθεί ότι, καταρχήν, τα κινησιογράμματα που δίνονται ως είσοδος στο εκάστοτε δίκτυο, έχουν αρκετές ομοιότητες οπτικά μεταξύ τους, καθώς πρόκειται για συγκεκριμένο είδος χορού, τον Ελληνικό Παραδοσιακό Χορό και, μάλιστα, από την ίδια περιφέρεια της Ελλάδας, την Περιφέρεια Ανατολικής Μακεδονίας και Θράκης. Οπότε, δεν είναι παράλογο το γεγονός ότι ο αλγόριθμος δεν μπορεί να επιτύχει ένα σαφή διαχωρισμό σε clusters. Επίσης, θα πρέπει να τονιστεί ότι τα scores αυτά επιτυγχάνονται χωρίς να έχει προηγηθεί εκπαίδευση των δικτύων με το συγκεκριμένο σύνολο δεδομένων, αλλά προ-εκπαίδευση σε εντελώς διαφορετικές εικόνες (ImageNet). Συνεπώς, η αποκτηθείσα γνώση μπορεί μεν να βοηθάει, αλλά από την άλλη δεν εμφανίζει ιδιαίτερος υψηλά ποσοστά επιτυχίας.

Για να κλείσει ομαλά το κεφάλαιο αυτό, καθώς και το Πρακτικό μέρος της εργασίας, θα ήθελα να κάνω ορισμένες παρατηρήσεις που αφορούν δοκιμές που έγιναν στα πλαίσια της εργασίας. Η πρώτη επισήμανση έχει να κάνει με το πλήθος των κινησιογραμμάτων και, συνεπώς, το πλήθος του συνόλου δεδομένων. Παρατηρήθηκε ότι όσο μικρότερος είναι ο αριθμός των δειγμάτων, τόσο υψηλότερα προκύπτουν τα τελικά αποτελέσματα και

μάλιστα στα δίκτυα όπου αξιοποιήθηκε η μεταφορά μάθησης παρατηρείται διαφορά ίση με 5-7% ανάλογα με το μοντέλο. Αυτό, ίσως να ερμηνεύεται, αν λάβουμε υπόψη μας την πολυπλοκότητα των συγκεκριμένων δειγμάτων, όπως επίσης και τη σχετική ομοιότητά τους, τα οποία δυσχεραίνουν τη διαδικασία της συσταδοποίησης. Οπότε, όσο λιγότερα τα δεδομένα, τόσο μικρότερη και η δυσκολία όσον αφορά το διαχωρισμό σε συστάδες. Η δεύτερη και τελευταία επισήμανση σχετίζεται με το πλήθος των συστάδων που επιλέγονται ως παράμετρος για τον αλγόριθμο k-means. Είδαμε ότι όσο μεγαλύτερος ήταν ο αριθμός των συστάδων, αντίστοιχα αυξανόταν και η επίδοση του clustering. Το γεγονός αυτό, ωστόσο, θεωρείται μάλλον λογικό, δεδομένου του ότι όσα περισσότερα τα clusters, τόσο πιο ομοιόμορφα θα είναι εντός, εφόσον θα περιλαμβάνουν λιγότερες κατηγορίες χορών το καθένα (και, μάλιστα, με μεγαλύτερη πιθανότητα να ανήκουν στο ίδιο είδος), και συγχρόνως διαφορετικά μεταξύ τους. Αυτή η αύξηση, όμως, του αριθμού των συστάδων οδηγεί, τελικά σε υπερ-κατάτμηση (*over-segmentation*), το οποίο δεν είναι θεμιτό.

## 5.5 Μεμονωμένες περιπτώσεις

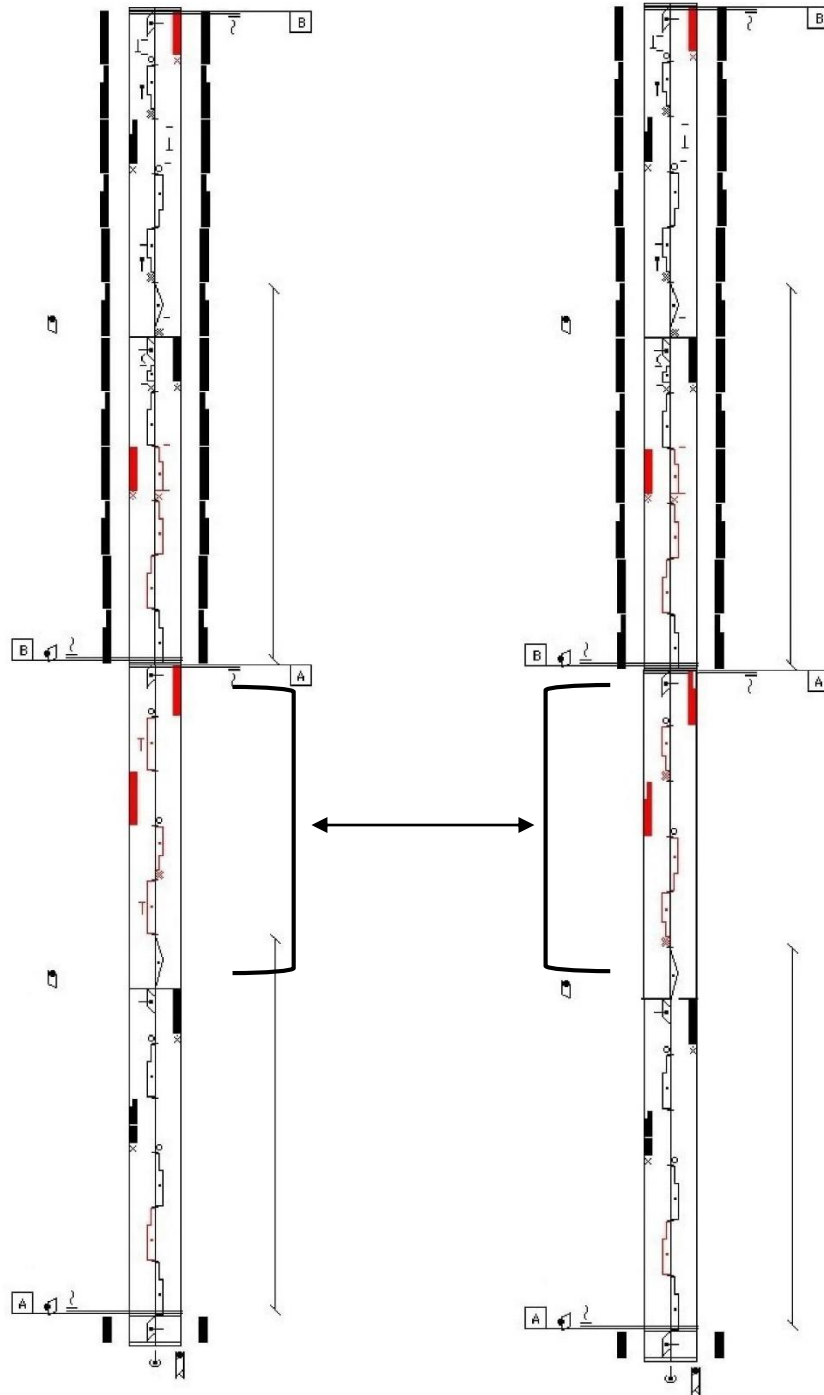
Στο τελευταίο υπο-κεφάλαιο της εργασίας, θελήσαμε να παρουσιάσουμε κάποια μεμονωμένα αποτελέσματα που θεωρούνται άξια αναφοράς και σχετίζονται είτε με κινησιογράμματα που πράγματι μοιάζουν ιδιαιτέρως, ανήκουν στην ίδια κατηγορία χορού και ομαδοποιούνται στην ίδια συστάδα, είτε με κινησιογράμματα που μοιάζουν οπτικά, αλλά δεν ανήκουν στην ίδια κατηγορία χορού, παρόλα αυτά αντιστοιχίζονται στην ίδια συστάδα. Όσα θα παρουσιάσουμε προέρχονται από τα αποτελέσματα που προέκυψαν, χρησιμοποιώντας το μοντέλο ResNet50 ως εξαγωγή χαρακτηριστικών.

### 5.5.1 Κινησιογράμματα από το ίδιο γένος Χορού

#### 5.5.1.1 Ο Γιάννης μας αρρώστησε

Αρχικά παραθέτω δύο κινησιογράμματα από τον χορό «*Ο Γιάννης μας αρρώστησε*», όπου βλέπουμε ότι οι παραλλαγές του αυτές διαφοροποιούνται σε ένα μόνο κινητικό μοτίβο, το 2<sup>ο</sup>, ενώ κατά τα άλλα είναι όμοιοι. Το ίδιο παρατηρείται και στις υπόλοιπες επτά παραλλαγές του, οπότε καταλαβαίνουμε ότι, πράγματι, παρουσιάζουν μεγάλη οπτική ομοιότητα. Ένα ιδιαίτερο χαρακτηριστικό αυτού του χορού είναι η ακολουθία στις λαβές των χεριών στο Β Μέρος, τις οποίες, προφανώς και αναγνώρισαν τα προ-εκπαιδευμένα μοντέλα, για αυτό και κατάφεραν να ομαδοποιήσουν όλες τις διαφορετικές εκδοχές αυτού

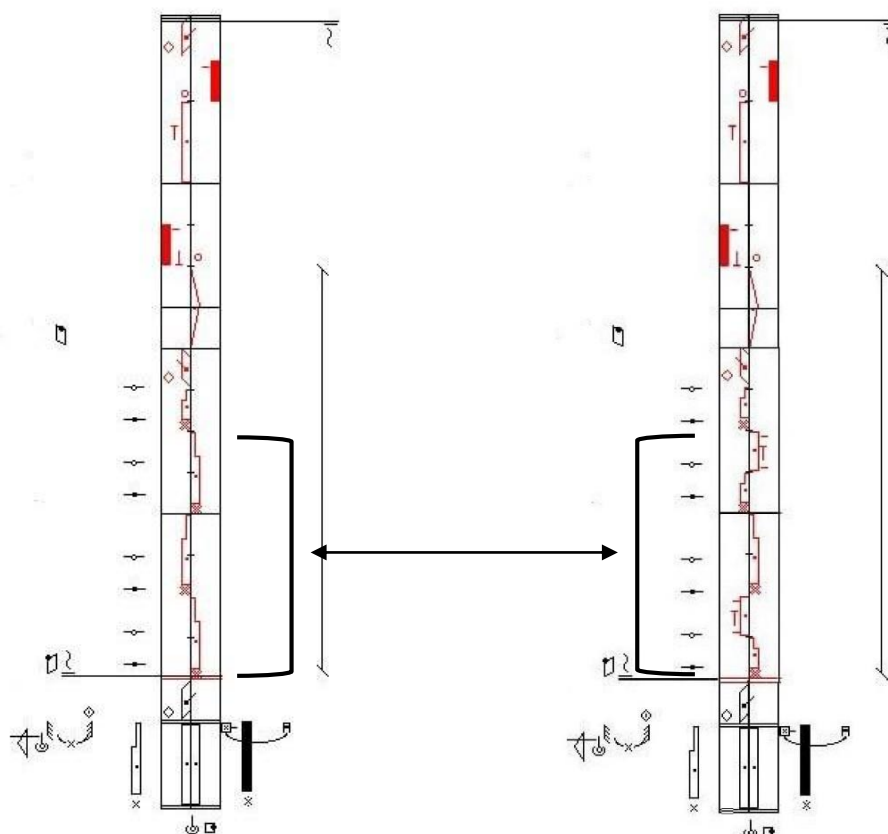
του χορού μαζί και, μάλιστα, ορισμένα εξ αυτών τον απομόνωσαν από διαφορετικά γένη χορών.



Σχήμα 5.18: Δύο παραλλαγές του χορού «Ο Γιάννης μας αρρώστησε».

### 5.5.1.2 Γκίκνα

Επίσης, ένα είδος χορού που τα μοντέλα κατάφεραν να ομαδοποιήσουν είναι η «Γκίκνα», μια υποκατηγορία του Ζωναράδικου. Και εδώ παρατηρούμε ότι η διαφορά τους έγκειται σε μερικά μόνο κινητικά στοιχεία. Ωστόσο, ο συγκεκριμένος χορός δεν παρουσιάζει κάποιο εξαιρετικά ιδιαίτερο χαρακτηριστικό γι' αυτό και πιθανώς τα μοντέλα τον αντιστοιχούν σε συστάδες όπου συναντώνται και άλλα υποείδη του Ζωναράδικου ή και άλλα γένη χορών γενικότερα.



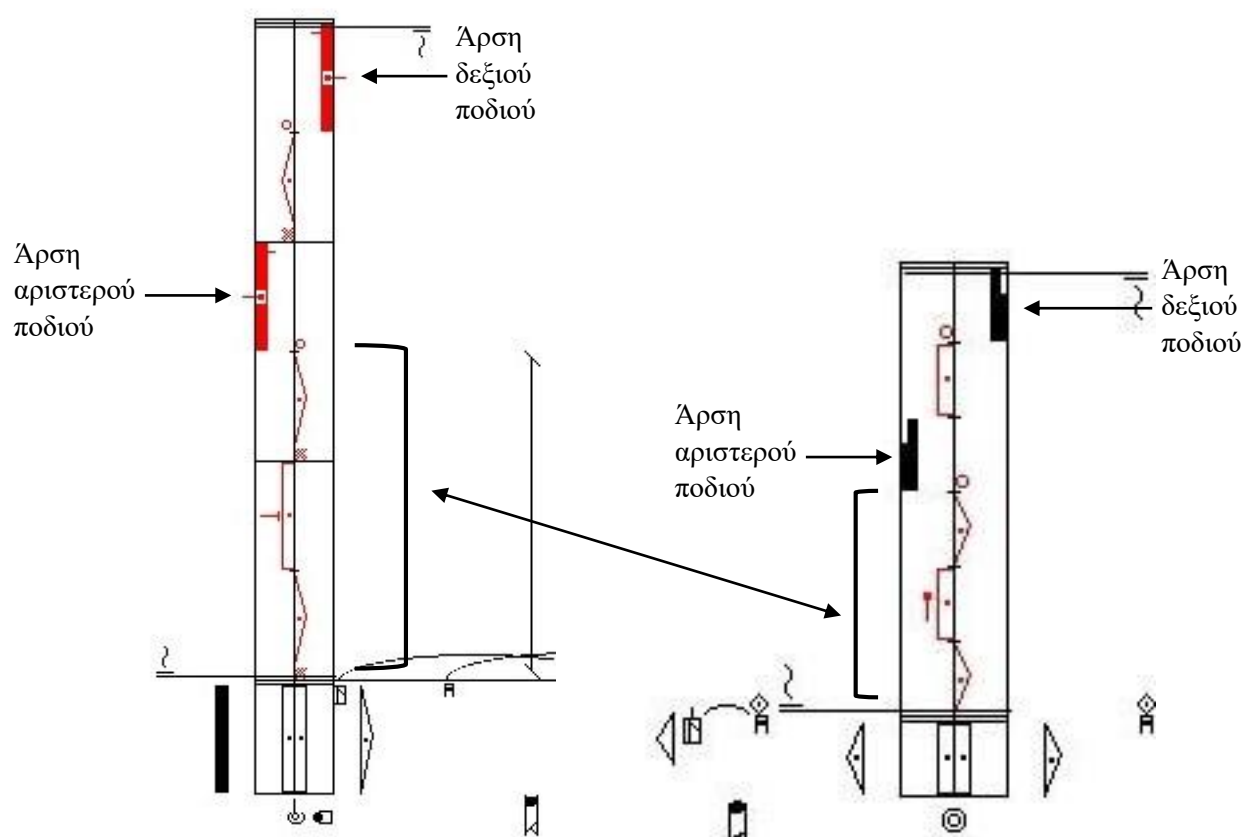
Σχήμα 5.19: Δύο παραλλαγές του χορού «Γκίκνα».

## 5.5.2 Κινησιογράμματα από διαφορετικά γένη Χορών

### 5.5.2.1 Ομοιότητα ως προς την ακολουθία κινήσεων

Πρόκειται για δύο κινησιογράμματα που ανήκουν στα γένη Χασάπικος και Ζωναράδικος αντίστοιχα, καθένα από τα οποία περιλαμβάνει 6 κινήσεις και, μάλιστα, οι 3 πρώτες ταυτίζονται, ενώ η 4<sup>η</sup> και η 6<sup>η</sup> παρότι διαφορετικές, πρόκειται και οι δύο για άρσεις

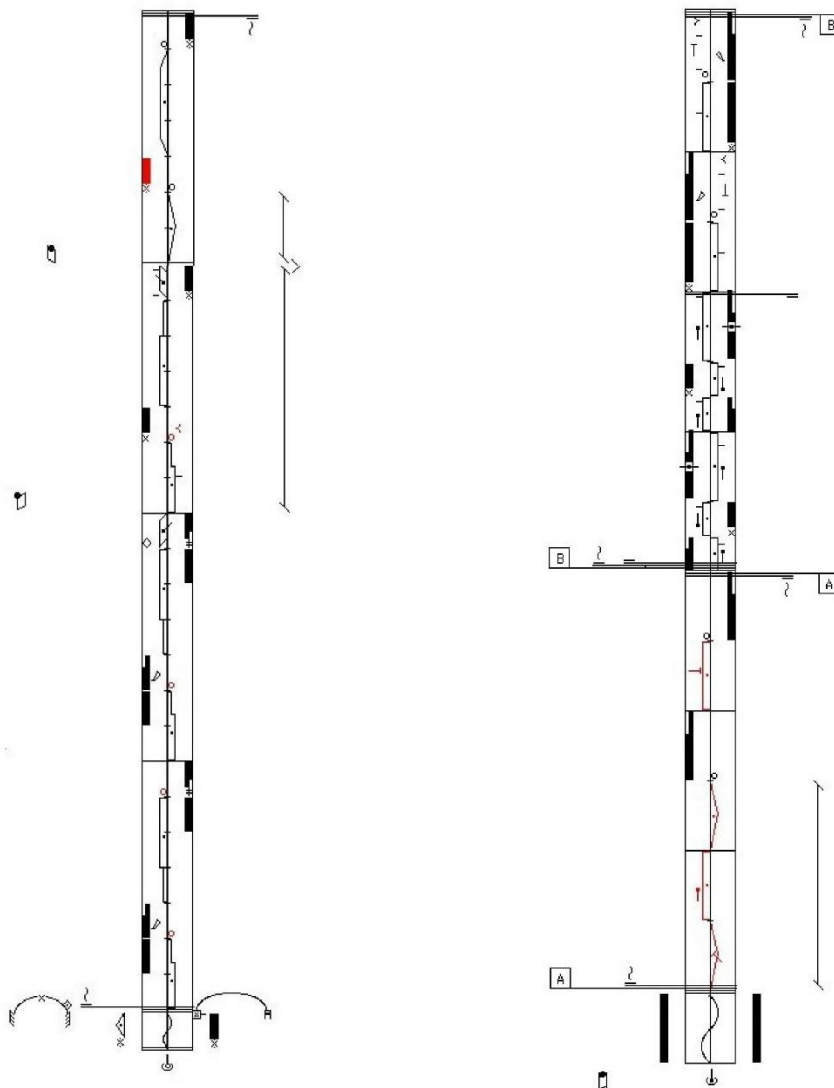
των ίδιων ποδιών. Επομένως, το μοντέλο αυτό, δικαίως, θα μπορούσαμε να πούμε ότι θεώρησε πιθανό να ανήκουν στην ίδια κατηγορία χορού.



Σχήμα 5.20: Κινησιογράμματα από τους χορούς Χασάπικος (αριστερά) και Ζωναράδικος (δεξιά).

### 5.5.2.2 Ομοιότητα ως προς τα επαναλαμβανόμενα μοτίβα

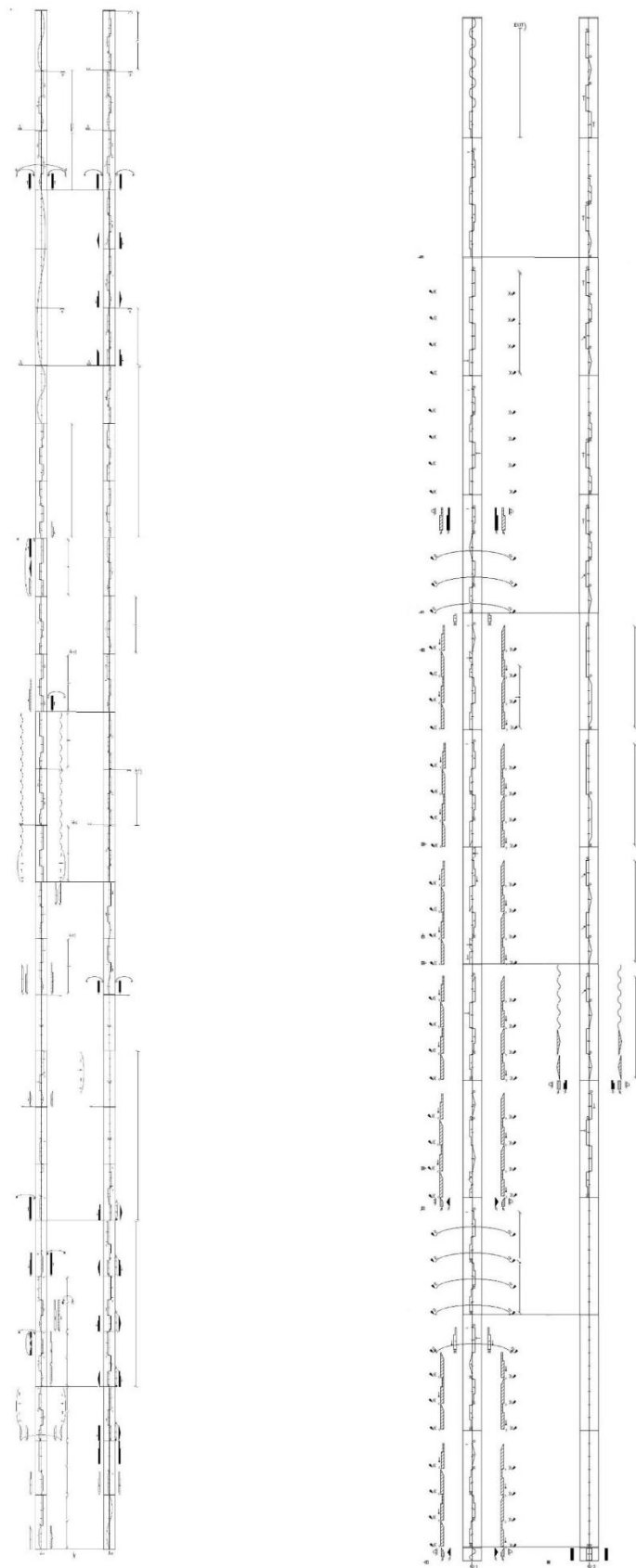
Πρόκειται για δύο χορούς τους «Κουκίτσα» και «Στης Ματζουράνας», καθένας από τους οποίους παρατηρούμε ότι περιλαμβάνει αρκετές άρσεις ποδιών (μαύρα γραμμοσκιασμένα σχήματα). Παρότι, δηλαδή, δε μοιάζουν ως προς τα βήματα, το μοντέλο ανακάλυψε αυτού του είδους το επαναλαμβανόμενο μοτίβο (άρσεις ποδιών) στα κινησιογράμματα για αυτό και τα ομαδοποίησε μαζί.



Σχήμα 5.21: Κινησιογράμματα από τους χορούς «Κουκίτσα» (αριστερά) και «Στης Ματζουράνας» (δεξιά).

### 5.5.2.3 Ομοιότητα ως προς το μήκος κινησιογράμματος

Εδώ παραθέτουμε δύο τελευταία κινησιογράμματα, όπου καθοριστικό ρόλο συντέλεσε το μήκος τους, δηλαδή το πλήθος των κινητικών μοτίβων – μουσικών μέτρων που περιλαμβάνουν. Είναι οι χοροί «Νταουλτζίδικος» (γένος Ζωναράδικος) και «Καρσιλαμάς» (παραλλαγή του). Πράγματι, πρόκειται για δύο εξαιρετικά εκτενή και πολύπλοκα κινησιογράμματα που, προφανώς, ομαδοποιήθηκαν στην ίδια συστάδα λόγω αυτού τους του χαρακτηριστικού. Μια άλλη ιδιαιτερότητα που αναγνωρίστηκε από το μοντέλο αυτό είναι ότι πρόκειται για χορούς που χορεύονται σε ζευγάρι.



Σχήμα 5.22: Κινησιογράμματα από τους χορούς «Νταουλιτζίδικος» (αριστερά) και παραλλαγή του Καρσιλαμά (δεξιά).





# Μέρος ΙΙΙ

Επίλογος

---



## Κεφάλαιο 6

### 6 Συμπεράσματα και Μελλοντικές Επεκτάσεις

---

Στο τελευταίο κεφάλαιο της παρούσας διπλωματικής εργασίας γίνεται μια σύνοψη της διαδικασίας που ακολουθήθηκε, αναδεικνύοντας τη σημασία και τα κύρια σημεία της. Ακόμη, εξάγονται ορισμένα βασικά συμπεράσματα, σαν απόρροια των πειραμάτων που διεξήχθησαν, επιτυγχάνοντας μια συνοπτική επισκόπηση των αποτελεσμάτων της εν λόγω έρευνας. Τέλος, προτείνονται ιδέες για πιθανές μελλοντικές επεκτάσεις που θα δώσουν την ευκαιρία εμβάθυνσης σε ερευνητικό επίπεδο σχετικά με το αντικείμενο που εξετάζεται στα πλαίσια της εργασίας.

#### 6.1 Σύνοψη

Στη μελέτη που προηγήθηκε εξετάζεται το πρόβλημα της Συσταδοποίησης κινήσιογραμμάτων, δηλαδή καταγεγραμμένων αλληλουχιών κίνησης που προέρχονται από τον Ελληνικό Παραδοσιακό Χορό και πιο συγκεκριμένα, περιλαμβάνουν 21 διαφορετικά είδη χορών της Ανατολικής Μακεδονίας και Θράκης. Αρχικά, γίνεται μια θεωρητική προσέγγιση επί του θέματος σχετικά με το χορό και την καταγραφή του, τονίζοντας τη σημασία ύπαρξης γραπτών τεκμηρίων, ιδίως στην περίπτωση του παραδοσιακού χορού που αποτελεί μέρος της άυλης πολιτιστικής κληρονομιάς ενός τόπου. Ακόμη, δίνεται το θεωρητικό υπόβαθρο που αφορά το τεχνικό κομμάτι της εργασίας, ώστε να γίνουν κατανοητές έννοιες, όπως αυτή της Μηχανικής Μάθησης, της Συσταδοποίησης, των Νευρωνικών Δικτύων, οι οποίες χρησιμοποιούνται κατά κόρον στο πρακτικό κομμάτι της προκείμενης εργασίας. Όλα αυτά, φυσικά, αποτελούν το κλειδί που θα χρησιμοποιηθεί για να ανοίξει την πύλη του πειραματικού μέρους του εγγράφου. Το τελευταίο εισάγεται με μια αναλυτική παρουσίαση του συνόλου δεδομένων και της προεπεξεργασίας που εκείνο υπέστη, ενώ, επίσης, γίνεται εκτενής αναφορά στην αρχιτεκτονική των δικτύων που χρησιμοποιήθηκαν. Στη συνέχεια, παρουσιάζονται τα πειράματα που διεξήχθησαν και ακολουθεί αξιολόγηση των αποτελεσμάτων τους, με διάφορες μετρικές. Σε αυτό το σημείο, θα ήθελα να εξηγήσω τον τρόπο εργασίας στα πειράματα που επιτελέστηκαν. Πρώτα από όλα, να σημειωθεί ότι το dataset, όπως αυτό μας δόθηκε από το Ερευνητικό Πρόγραμμα «Θράκη – Μακεδονία», δε χρησιμοποιήθηκε αυτούσιο, αλλά υπέστη επεξεργασία, με σκοπό να είναι εφικτή η χρήση του στο πρακτικό μέρος της έρευνάς μας. Ακολούθησε

επιπλέον προεπεξεργασία των δεδομένων, σε τεχνικό αυτή τη φορά επίπεδο, χρησιμοποιώντας τεχνικές Βαθιάς Μηχανικής Μάθησης για την εξαγωγή των χρήσιμων χαρακτηριστικών τους, καθώς και μεθόδους μείωσης της διαστατικότητάς τους. Εν συνεχεία, τα διανύσματα που προέκυψαν, χρησιμοποιήθηκαν σαν είσοδος στον αλγόριθμο συσταδοποίησης, ο οποίος παρήγαγε και τα τελικά αποτελέσματα. Το πρακτικό μέρος της εργασίας ολοκληρώνεται αισίως με την παρουσίαση και αξιολόγηση αυτών των αποτελεσμάτων, για την οποία επιλέχθηκαν τέσσερις διαφορετικές μετρικές που ελέγχουν την εγκυρότητα μιας συσταδοποίησης.

## 6.2 Συμπεράσματα

Η συγκεκριμένη μελέτη έγινε στα πλαίσια μιας προσπάθειας να συνδυαστούν δύο ενδιαφέροντες επιστημονικοί κλάδοι, αυτός της Μηχανικής Μάθησης με εκείνον του Χορού, που μόνο τα τελευταία χρόνια φαίνεται να έχουν ξεκινήσει να αλληλεπιδρούν. Σε πρώτη φάση, πραγματοποιήθηκε βιβλιογραφική ανασκόπηση, με στόχο την εύρεση εργασιών και ερευνών που θα μπορούσαν να συμβάλλουν στην επιλογή κατευθύνσεων δράσης. Δυστυχώς, τα ευρήματα δεν ήταν ικανοποιητικά και αυτό αποδίδεται σε μεγάλο βαθμό στην έλλειψη επαρκών δεδομένων που να σχετίζονται με τον τομέα της σημειογραφίας του Χορού και, ακόμη περισσότερο, με αυτόν της καταγραφής παραδοσιακών χορών. Ένας ακόμη λόγος που δε βρέθηκε ο προσδοκώμενος όγκος πληροφοριών (αν και αναμενόμενο), είναι η πολύ πρόσφατη εκδήλωση ενδιαφέροντος από μεριάς της επιστημονικής κοινότητας για θεματικές που αφορούν τις παραστατικές τέχνες και ιδίως το χορό, λόγω της εγγενούς πολυπλοκότητάς του και της συχνής δυσκολίας του ως προς την ερμηνεία. Όλα αυτά οδήγησαν στη συστηματική μελέτη εργασιών που είχαν, βέβαια, σαν βάση την καταγραφή της χορευτικής κίνησης, αλλά εστίαζαν περισσότερο στην αυτόματη παραγωγή κινησιογραμμάτων, παρά στην αναγνώριση, ταξινόμηση ή συσταδοποίηση ήδη υπαρχόντων παρτιτούρων.

Δεδομένης, ωστόσο, της ύπαρξης ενός εξαιρετικά σημαντικού και δύσκολου ως προς τη συλλογή, συνόλου δεδομένων, το οποίο ήταν ανοιχτό στο ευρύ κοινό, αποφασίστηκε αυτό να αξιοποιηθεί και, μάλιστα, με τρόπο που δεν είχε αναλυθεί ξανά στο παρελθόν. Πρόκειται για ένα σύνολο 271 εικόνων που περιλαμβάνουν καταγεγραμμένα χορευτικά μοτίβα από παραδοσιακούς χορούς της Θράκης, ένα γραπτό τεκμήριο, κειμήλιο του ελληνισμού. Έτσι, προέκυψε η ιδέα της συσταδοποίησης των εν λόγω κινησιογραμμάτων, με σκοπό να προβούμε σε μια πρώτη επαφή με αυτό το σύνολο δεδομένων, κάτι που θα βοηθήσει στην περαιτέρω ανάλυσή του και την εξαγωγή χρήσιμης γνώσης. Λόγω του ότι, το dataset αποτελούνταν από εικόνες, καταλήξαμε στο ότι θα ήταν χρήσιμη μια οπτική

σύγκρισή τους, για την αναγνώριση μοτίβων και προτύπων που θα τις κατέτασσαν, τελικά, σε ίδιες ή διαφορετικές συστάδες. Επομένως, η εν λόγω διπλωματική μελέτησε τη συσταδοποίηση αυτών των εικόνων με βάση την οπτική τους ομοιότητα και αυτός ήταν ο λόγος, για τον οποίο επιλέχθηκε η εμβάθυνση στο πρόβλημα με τη χρήση νευρωνικών δικτύων, τα οποία είναι ιδιαίτερος αποτελεσματικά στην εξαγωγή ισχυρών χωρικών τοπικών συσχετισμών στα δεδομένα.

Ας διερευνήσουμε, ωστόσο, τη διαδικασία εκπόνησης της παρούσας διπλωματικής εργασίας από την αρχή, εκφράζοντας και τα αντίστοιχα συμπεράσματα. Πρώτα από όλα, λοιπόν, εξετάσαμε τα δεδομένα, τα οποία αν και γραμμένα με ομοιόμορφο τρόπο, σε ψηφιακή μορφή, γεγονός που καθιστά ευκολότερη την επεξεργασία τους από ένα σύστημα μηχανικής μάθησης, παρουσίαζαν το μειονέκτημα ότι θα έπρεπε να υποστούν προ-επεξεργασία «με το χέρι», έτσι ώστε να μπορέσουν να χρησιμοποιηθούν ως είσοδος σε ένα σύστημα που έχουμε σαν τελικό στόχο να βγάλει όσο το δυνατόν πιο εύστοχα αποτελέσματα. Το κυριότερο εμπόδιο που αντιμετωπίσαμε ήταν η τμηματική παρουσίαση των παραλλαγών κάθε χορού, δηλαδή αντί να δίνεται το κινησιόγραμμα της κάθε παραλλαγής, δίνονταν μόνο το βασικό βηματολόγιο του εκάστοτε χορού και ως διαφορετικές εκδοχές του παρουσιάζονταν μόνο τα τροποποιημένα κινητικά μοτίβα του, με άλλα λόγια ορισμένα μόνο κομμάτια των κινησιογραμμάτων. Φυσικά, εάν ένα ανθρώπινο μάτι μελετά τις συγκεκριμένες παρτιτούρες, είναι σχετικά εύκολο να κατανοήσει ότι αυτά τα κομμάτια αποτελούν μέρος του όλου, που αντικαθιστούν συγκεκριμένα κινητικά μοτίβα από τον «αρχικό χορό». Για να γίνει, ωστόσο, μια σωστή οπτική σύγκριση, με τη βοήθεια τεχνικών μηχανικής μάθησης που ήταν και ο απώτερος σκοπός της εργασίας, αυτό δεν είναι προφανές και αλλοιώνει την ποιότητα των δεδομένων και κατ' επέκταση των αποτελεσμάτων. Αυτό είχε σαν συνέπεια, οι παρτιτούρες αυτές να μην μπορούν να χρησιμοποιηθούν αυτούσιες και επομένως, να καθίσταται αναγκαίο «να κοφτούν και να ραφτούν» από εμάς. Συμπερασματικά, θα ήταν ιδιαίτερος βοηθητικό εάν οι ειδικοί που ασχολούνται με τη συλλογή και καταγραφή των δεδομένων του χορού, φρόντιζαν για την ολοκληρωμένη παρουσίαση των παρτιτούρων, και όχι επιμέρους μόνο τμημάτων τους, ώστε να είναι, στη συνέχεια, αποτελεσματική η επεξεργασία τους από συστήματα αναγνώρισης προτύπων, μηχανικής μάθησης ή και όρασης υπολογιστών.

Αμέσως επόμενη διαπίστωση, στην οποία καταλήξαμε εξετάζοντας το τελικό (ελαφρώς τροποποιημένο σε σχέση με το αρχικό) σύνολο δεδομένων μας, είναι ότι παρουσιάζει σε μεγάλο βαθμό ανισορροπία, πρόκειται δηλαδή για ένα imbalanced dataset. Το γεγονός αυτό, προκαλεί δυσκολίες εξ' ορισμού και γνωρίζαμε εκ των προτέρων ότι θα επηρέαζε αρνητικά τις τεχνικές που χρησιμοποιούνται στο πρακτικό μέρος της εργασίας

και σα φυσικό επακόλουθο, τα τελικά αποτελέσματα. Η ανισορροπία αυτή εξηγείται ως ένα βαθμό, δεδομένου του ότι υπήρχε πολύ μεγάλη συγκέντρωση κινησιογραμμάτων του Ζωναράδικου, Χασάπικου, Συγκαθιστού και Καρσιλαμά, δηλαδή από δημοφιλή γένη χορών, που περιλαμβάνουν αρκετά υποείδη ο καθένας και χορεύονται στην πλειοψηφία των χωριών από όπου έγινε η συλλογή των δεδομένων. Οπότε, δεν ήταν άλογο το ότι προέκυψε ανομοιομορφία στο πλήθος των δειγμάτων κάθε γένους χορού. Ωστόσο, στα πλαίσια της συγκεκριμένης εργασίας επιλέχθηκε να μην χρησιμοποιηθούν τεχνικές διαχείρισης αυτής της ανισορροπίας, όπως είναι οι τεχνικές επαναδειγματοληψίας (resampling techniques) ή επαύξησης δεδομένων (data augmentation), καθώς ως διεπιστημονική μελέτη που σχετίζεται με μια θεματική, η οποία δεν έχει πρωτότερα ερευνηθεί, προτιμήσαμε να δουλέψουμε με τα πρωτότυπα δεδομένα, στην παρθενική της προσέγγιση τουλάχιστον, χωρίς οποιοδήποτε είδους αλλοίωση αυτών.

Παρόλα αυτά, να σημειωθεί το γεγονός ότι έγιναν αρκετές δοκιμές που σχετίζονται με σύγχρονες μεθόδους προεπεξεργασίας εικόνων, όμως καμία δε συνείσφερε στη βελτίωση των τελικών αποτελεσμάτων, για αυτό και περιοριστήκαμε σε τρεις μόνο τεχνικές: στην αλλαγή και κατάλληλη προσαρμογή του μεγέθους των εικόνων, αναλόγως με το κάθε μοντέλο, την εξαγωγή χρήσιμων χαρακτηριστικών που θα έδιναν μια πιο σαφή «εικόνα» σχετικά με την υποκείμενη δομή (underlying structure) των δεδομένων και, τέλος, την μείωση της διαστατικότητάς τους. Όσον αφορά την τελευταία, ασχοληθήκαμε αρκετά και δώσαμε έμφαση στη βελτιστοποίηση των υπερπαραμέτρων της, παρότι σε κάθε περίπτωση που μελετήθηκε προέκυπτε διαχειρίσιμο πλήθος χαρακτηριστικών, οπότε δεν ήταν απαραίτητη η εφαρμογή μιας τέτοιας τεχνικής. Παρόλα αυτά, τη θεωρήσαμε σημαντική, πρώτον γιατί βοηθά στην επιλογή της χρήσιμης πληροφορίας που πρέπει να διατηρηθεί, απορρίπτοντας, συχνά «θορυβώδη» δεδομένα και, επιπροσθέτως, συμβάλλει στην όποια εξοικονόμηση υπολογιστικής ισχύος. Πράγματι, αυτή η μείωση των διαστάσεων αποδείχτηκε ιδιαίτερος βοηθητική στο κομμάτι της συσταδοποίησης, κατά την προσπάθεια εύρεσης των κατάλληλων υπερπαραμέτρων, όπου ο κώδικας χρειάστηκε να εκτελεστεί πάρα πολλές φορές, το οποίο ήταν αρκετά χρονοβόρο, οπότε εάν δουλεύαμε με το αρχικό πλήθος χαρακτηριστικών οι χρονικές επιδόσεις θα ήταν απογοητευτικές.

Το επόμενο βήμα, ήταν αυτό που καθόρισε σε μεγάλο βαθμό τα αποτελέσματα της εργασίας, για αυτό και δόθηκε ιδιαίτερη προσοχή και πραγματοποιήθηκαν πολυάριθμες δοκιμές. Αναφερόμαστε στην επιλογή των κατάλληλων νευρωνικών δικτύων, η οποία δεν ήταν κάτι το προφανές και προκαθορισμένο, ενώ η έλλειψη παρόμοιων ερευνών και προτάσεων που θα μπορούσαν ενδεχομένως να λειτουργήσουν σαν βάση, έκανε ελαφρώς πιο δύσκολη την έρευνά μας. Μια ασφαλής επιλογή, ωστόσο, ήταν αδιαμφισβήτητα τα

Συνελκτικά Νευρωνικά Δίκτυα, δεδομένης της υψηλής απόδοσής τους σε δεδομένα μορφής εικόνας. Έτσι, χρησιμοποιήθηκαν, ώστε να μετασχηματίσουν τα δείγματα – εικόνες σε αναπαραστάσεις πιο χρήσιμες και «καθαρές» για τον εκάστοτε αλγόριθμο συσταδοποίησης. Πιο συγκεκριμένα, για τη δημιουργία μοντέλων που θα εξυπηρετούσαν αποδοτικά τον παραπάνω σκοπό, ακολουθήθηκε η μεθοδολογία της «σταδιακής εξέλιξης», δηλαδή ξεκινώντας από το πιο απλό συνελκτικό δίκτυο, συνεχίσαμε υλοποιώντας πιο περίπλοκες αρχιτεκτονικές, καταλήγοντας να κάνουμε μεταφορά γνώσης από σύγχρονα, βαθιά μοντέλα. Στόχος ήταν η επισκόπηση της συμπεριφοράς τους και της χρησιμότητας των χαρακτηριστικών που το εκάστοτε δίκτυο εξάγει, ώστε να επιλεγεί το αποδοτικότερο για το προκείμενο πρόβλημα. Κάτι το οποίο παρατηρήσαμε, στα πρώτα τρία μοντέλα, σε εκείνα, δηλαδή, που εμείς δημιουργήσαμε, είναι ότι τα επίπεδα Batch Normalization (κανονικοποίησης δέσμης), δημιούργησαν αισθητή διαφορά στα αποτελέσματα. Επιπλέον, αν και προφανές, διαπιστώσαμε ότι όσο πιο βαθύ το δίκτυο, τόσο πιο χρήσιμα χαρακτηριστικά εξάγονται από τα δεδομένα και, εν τέλει, τόσο καλύτερη και η απόδοση του συστήματος. Αυτό αποδείχθηκε και από τα μη, αλλά και από τα προ-εκπαιδευμένα μοντέλα που χρησιμοποιήθηκαν.

Κλείνοντας, θα ήθελα να αναφερθώ στη διαδικασία της συσταδοποίησης, όπου εστίασαμε στον αλγόριθμο k-means και τη βελτιστοποίηση των υπερπαραμέτρων του, για να διερευνηθεί η συμπεριφορά του ως προς τα συγκεκριμένα δεδομένα, αλλά και τα συγκεκριμένα χαρακτηριστικά που εξήγαγαν τα συνελκτικά μοντέλα που χρησιμοποιήθηκαν. Παρατηρήσαμε ότι πρόκειται για έναν απλό και ταχύ αλγόριθμο, εύκολα διαχειρίσιμο που, ωστόσο, δεν παρουσίασε υψηλές επιδόσεις στο πρόβλημά μας. Συνεργάστηκε καλύτερα με τα μοντέλα ResNet50 και InceptionV3, (όπως, πιθανώς, και άλλοι αλγόριθμοι συσταδοποίησης), δίνοντας σχετικά αξιοπρεπή αποτελέσματα, χωρίς υψηλά ποσοστά επιτυχίας. Δε χρησιμοποιήθηκαν εναλλακτικοί αλγόριθμοι συσταδοποίησης, καθώς η αλήθεια είναι ότι, προτιμήθηκε η εμβάθυνση στην εξαγωγή χαρακτηριστικών, παρά στη συσταδοποίηση αυτή κάθε αυτή. Έτσι, δόθηκε μεγαλύτερη βαρύτητα στη χρήσιμη πληροφορία, δηλαδή στα χαρακτηριστικά που μπορούν να εξαχθούν από το σύνολο δεδομένων, καθώς αυτά είναι που θα καθορίσουν, τελικά, και τα αποτελέσματα. Όσο πιο «καθαρά» αναπαριστούν τα κρυμμένα μοτίβα των εικόνων, τόσο μεγαλύτερη θα είναι και η απόδοση οποιουδήποτε αλγορίθμου. Για αυτό το λόγο, αποφασίστηκε, σε πρώτη φάση, να υλοποιηθεί εκτενής μελέτη ως προς τα κατάλληλα νευρωνικά μοντέλα και, σε μελλοντικό στάδιο, να βρεθεί και ο κατάλληλος αλγόριθμος συσταδοποίησης, βάσει της φύσης των δειγμάτων.

Συνολικά, θεωρώ κρίσιμο να αναφερθεί ότι τα παραγόμενα αποτελέσματα που αξιολογούν την ποιότητα του clustering για το συγκεκριμένο πρόβλημα, σίγουρα δεν φτάνουν το ποσοστό επιτυχίας που μπορεί να τραβούσε την προσοχή του αναγνώστη, όπως συμβαίνει σε άλλα προβλήματα Βαθιάς Μάθησης. Ο στόχος της εργασίας, όμως, δεν ήταν αυτός εξαρχής. Μέσα από αυτή τη μελέτη, θελήσαμε να προσεγγίσουμε ένα ενδιαφέρον πρόβλημα που συνδυάζει στοιχεία από δύο επιστήμες, για να ανακαλύψουμε πώς το σύστημα που δημιουργήθηκε θα διαχειριστεί αυτήν την πολύ εξειδικευμένη πληροφορία που του παρέχεται και με την οποία δεν είναι εξοικειωμένο, ώστε να καταλήξουμε σε συμπεράσματα που θα μας βοηθήσουν να κάνουμε βήματα προς τα εμπρός, ευελπιστώντας ότι θα ανοίξουν τις πύλες σε νέα ερευνητικά πεδία. Επομένως, το θεωρώ άδικο, να πει κανείς ότι το πρόβλημα αυτό δεν είχε μεγάλη επιτυχία, αντιθέτως θα έλεγα πως μόλις μπήκαν τα θεμέλια για κάτι ενδιαφέρον που μπορεί να διευκολύνει σημαντικά και να αναδείξει τη σπουδαιότητα του κλάδου της Σημειογραφίας του Χορού.

### 6.3 Μελλοντικές επεκτάσεις

Χρησιμοποιώντας την εν λόγω εργασία και τα συμπεράσματα που εξήχθησαν, σαν σημείο αφετηρίας, μπορεί να προκύψει πληθώρα προτάσεων, όσον αφορά τη βελτιστοποίηση των αποτελεσμάτων της συγκεκριμένης μελέτης, αλλά, φυσικά, και να δοθούν ιδέες για πιθανές τροποποιήσεις και νέες προοπτικές για περαιτέρω έρευνα.

Αρχικά, όσον αφορά το σύνολο δεδομένων, το οποίο δεν ήταν ιδανικό για επεξεργασία μέσω τεχνικών Βαθιάς Μάθησης, όπως, δυστυχώς, συμβαίνει συχνά στα ευάριθμα δεδομένα που προέρχονται από το χώρο του Χορού, θα ήταν εξαιρετικά χρήσιμο, εάν κάποιος/α σημειογράφος ασχολούνταν με την εκ νέου γραφή αυτών των κινησιογραμμάτων, με τον τρόπο που προτείνεται στην παραπάνω υποενότητα, δηλαδή κάθε χορός, αλλά και κάθε παραλλαγή του να έχει τη δική του παρτιτούρα, με τα νέα κινησιογράμματα να είναι όσο το δυνατόν πιο ευκρινή. Επιπροσθέτως, από τεχνικής πλευράς, θα μπορούσαν να διερευνηθούν εναλλακτικοί τρόποι προεπεξεργασίας των εικόνων και να γίνει εκτενέστερη μελέτη σχετικά με την κατάλληλη επιλογή των παραμέτρων τους, ώστε να εξεταστεί εάν θα μπορούσε αυτό να βελτιώσει τα παραγόμενα αποτελέσματα.

Μια ακόμη πρόταση που αφορά το σύνολο δεδομένων, είναι το να επιχειρηθούν ορισμένες τεχνικές διαχείρισης της ανισορροπίας που αυτό παρουσιάζει, λόγω του ότι αποτελεί ένα σημαντικό παράγοντα που επηρεάζει αρνητικά την απόδοση των συστημάτων που αναπτύχθηκαν. Χρήσιμες, πιθανώς, να αποδεικνύονταν ορισμένες τεχνικές επαναδειγματοληψίας, γνωστές ως Resampling techniques, οι οποίες τροποποιούν



τα δεδομένα για να εξισορροπήσουν το διαχωρισμό τους στις κλάσεις μέσω του oversampling (υπερδειγματοληψία) και του undersampling (υποδειγματοληψία). Επιπλέον, ορισμένες ευρέως χρησιμοποιούμενες μέθοδοι για την επίλυση τέτοιων προβλημάτων, είναι εκείνες που κάνουν επαύξηση δεδομένων, Data Augmentation, οι οποίες χρησιμοποιώντας διάφορους μετασχηματισμούς επί των αρχικών δεδομένων, δημιουργούν νέα τροποποιημένα δείγματα, με σκοπό να εμπλουτίσουν τις minority classes.

Επίσης, θα μπορούσαμε να εμβαθύνουμε περαιτέρω σε Συνελκτικά Νευρωνικά Δίκτυα που δε χρησιμοποιήθηκαν στα πλαίσια της εργασίας, μέσω της Μεταφοράς Μάθησης. Ακριβώς, επειδή, χρησιμοποιούνται σαν εξαγωγείς χαρακτηριστικών για το σύνολο δεδομένων μας, είναι σημαντικό μέσα από την αστείρευτη πηγή μοντέλων που έχουν, πλέον, αναπτυχθεί, να χρησιμοποιηθεί εκείνο που θα εξάγει την «πιο χρήσιμη» πληροφορία. Ενδιαφέρον, ακόμη, θα ήταν να διερευνηθεί η περίπτωση δημιουργίας υβριδικών μοντέλων, δηλαδή αρχιτεκτονικών που έχουν προκύψει από το συνδυασμό δύο ή περισσότερων επιμέρους νευρωνικών δικτύων με στόχο την αξιοποίηση των καλύτερων στοιχείων τους, δημιουργώντας ισχυρότερα μοντέλα. Με αυτόν τον τρόπο, θα είναι, θεωρητικά, ευκολότερη και αποτελεσματικότερη η ανακάλυψη των κρυμμένων μοτίβων και των εξειδικευμένων λεπτομερειών των κινησιογραμμάτων.

Μια ακόμη κατεύθυνση για μελλοντική ανάπτυξη, θα μπορούσε να αποτελέσει η διερεύνηση διαφορετικών αλγορίθμων συσταδοποίησης. Σε πρώτη φάση, θεωρώ σημαντική τη δοκιμή τεχνικών από διαφορετικά είδη clustering, όπως είναι οι ιεραρχικοί αλγόριθμοι (Hierarchical Clustering), εκείνοι με βάση την πυκνότητα (Density-based), τα πλέγματα (Grid-based) ή τα μοντέλα (Model-based), δηλαδή να επιλεγούν ευρέως χρησιμοποιούμενες μέθοδοι που αντιπροσωπεύουν διαφορετικές κατηγορίες της συσταδοποίησης. Κύριος στόχος είναι να βρεθούν αλγόριθμοι πιο «ανθεκτικοί» σε imbalanced data, ώστε τα τελικά αποτελέσματα να μην επηρεάζονται σε μεγάλο βαθμό από το σημαντικό αυτό μειονέκτημα του συνόλου δεδομένων. Σε επόμενο στάδιο, αφού επιλεγεί η κατηγορία που ταιριάζει περισσότερο στο εν λόγω πρόβλημα, θα μπορούσε να γίνει περαιτέρω εμβάθυνση, εξετάζοντας τους επιμέρους αλγορίθμους που εκείνη περιλαμβάνει.

Επιπλέον, όσον αφορά το κομμάτι της συσταδοποίησης, θα ήταν ενδιαφέρον να μελετηθεί ο διαχωρισμός των δεδομένων σε συστάδες, χρησιμοποιώντας σαν βάση, άλλα στοιχεία που θα μπορούσαμε να συλλέξουμε από τα μεταδεδομένα κάθε χορού, όπως είναι για παράδειγμα το υποείδος του. (Στην περίπτωση αυτή, όσοι χοροί δεν περιλαμβάνουν κάποιο υποείδος, θα χρησιμοποιούνταν σαν ετικέτα, η βασική τους κατηγορία.) Μια ακόμη ενδιαφέρουσα βάση για ομαδοποίηση, θα μπορούσε να αποτελεί η περιοχή από όπου

προέρχεται ο καθένας. Θα ήταν χρήσιμο, μάλιστα, να παρουσιαστούν συγκριτικά τα αποτελέσματα των συσταδοποιήσεων γένους και περιοχής αντιστοίχως, ώστε να παρατηρήσουμε ομοιότητες και διαφορές και να εξάγουμε συμπεράσματα σχετικά με το τι, τελικά, διαδραματίζει σημαντικότερο ρόλο, ο τόπος από όπου προέρχεται κάποιος χορός, ή η κατηγορία στην οποία ανήκει. Γιατί, ας μην ξεχνάμε ότι στην Ελλάδα υπάρχουν αρκετοί παραδοσιακοί χοροί, οι οποίοι αν και έχουν το ίδιο όνομα χορεύονται με σημαντικά διαφορετικό τρόπο, ανάλογα με την περιοχή, όπου δημιουργήθηκαν.

Θα ήθελα, επίσης, να παρουσιάσω μια διαφορετική προοπτική για το προκείμενο πρόβλημα, στην οποία κατέληξα μέσα από τις πολυάριθμες προσπάθειες που έγιναν για τη βελτιστοποίηση των αποτελεσμάτων. Λόγω της φύσης των συγκεκριμένων δεδομένων και της πολυπλοκότητας που παρουσιάζουν, θα ήταν, θεωρώ, χρησιμότερο προτού κατευθυνθούμε προς τη μη Επιβλεπόμενη Μάθηση, όπως είναι η συσταδοποίηση που μελετάται στην εν λόγω εργασία, να αξιοποιήσουμε τα οφέλη της Επιβλεπόμενης Μάθησης. Με αυτό εννοώ ότι θα ήταν, ενδεχομένως, χρησιμότερο εάν, πρώτα από όλα, εκπαιδεύαμε στο συγκεκριμένο σύνολο δεδομένων, τα δίκτυα που χρησιμοποιούμε, καθώς εξαιτίας της τόσο μεγάλης απόκλισης των κινησιογραμμάτων σε σχέση με τις εικόνες, στις οποίες έχουν προ-εκπαιδευτεί ορισμένα μοντέλα, όπως είναι το ImageNet, εν τέλει, τα δίκτυα αυτά δεν εξάγουν με μεγάλη επιτυχία τα σημαντικά εκείνα χαρακτηριστικά των παρτιτούρων χορού, και κατ' επέκταση δεν επιτυγχάνονται υψηλές επιδόσεις στο πρόβλημα της συσταδοποίησης. Έτσι, μια πρόταση που θα έκανα για μελλοντική υλοποίηση, δεδομένου του ότι, για τις ανάγκες αυτής της εργασίας έχει ήδη δημιουργηθεί αρχείο με τις ετικέτες του κάθε δείγματος, είναι η ανάπτυξη ενός συστήματος Επιβλεπόμενης Μάθησης. Σε ένα πρώτο στάδιο, θα μπορούσε να εκτελεί ταξινόμηση των χορών με βάση το γένος τους, οπότε θα κατέτασσε τα κινησιογράμματα σε μία από τις 21 διαφορετικές κλάσεις, το οποίο, φυσικά, ενέχει αρκετά μεγάλο βαθμό πολυπλοκότητας, αλλά πιστεύω ότι είναι πολλά υποσχόμενο. Σε επόμενο στάδιο, τα μοντέλα αυτά, όντας πια κατάλληλα προ-εκπαιδευμένα, θα μπορούσαν να χρησιμοποιηθούν σαν εξαγωγείς χαρακτηριστικών για τη μετέπειτα συσταδοποίηση του συνόλου δεδομένων, κάτι το οποίο, σίγουρα θα βελτιώνει αισθητά τα τελικά αποτελέσματα.

Τέλος, μια ενδιαφέρουσα και ιδιαίτερος χρήσιμη πρόταση για το μέλλον, που έπεται των παραπάνω βελτιστοποιήσεων, καθώς προϋποθέτει υψηλότερη απόδοση στο εν λόγω πρόβλημα, σχετίζεται με την ανάπτυξη εφαρμογών (applications) που θα μπορούσαν να χρησιμοποιήσουν σα βάση και να εξελίξουν το σύστημα που προτείνεται στη διπλωματική αυτή εργασία. Καταρχήν, θα μπορούσε να υλοποιηθεί ένα application, το οποίο να προσφέρει είτε συσταδοποίηση των κινησιογραμμάτων με βάση το γένος των χορών όπου

ανήκουν, είτε ακόμη και αναγνώριση του είδους τους. Μέσω μιας εύχρηστης διεπαφής χρήστη (user interface), θα μπορούσε το εκάστοτε άτομο να παρέχει στην εφαρμογή ένα κινησιόγραμμα και να ενημερώνεται από αυτήν, σχετικά με το είδος στο οποίο εντάσσεται ή να μαθαίνει και άλλες σχετικές με το χορό πληροφορίες που τον ενδιαφέρουν. Επιπροσθέτως, μια εναλλακτική «έξυπνη» προσέγγιση θα ήταν η ανάπτυξη μιας εφαρμογής που ζητά από τον ίδιο το χρήστη να ταξινομήσει ορισμένα κινησιογράμματα στις κατηγορίες που ανήκουν. Αυτό θα μπορούσε να αποτελέσει ένα ενδιαφέρον παιχνίδι, για παιδιά, εφήβους και ενήλικες που έχουν σχετικά ενδιαφέροντα, φέρνοντάς τους σε επαφή όχι μόνο με τις ρίζες τους, αλλά και με τη σημειογραφία του χορού. Φυσικά, εφόσον τα συστήματα καταγραφής της χορευτικής κίνησης, όπως έχει ήδη εξηγηθεί, είναι σύνθετα ως προς την ανάγνωση και απαιτούν εμπειρία, η εφαρμογή αυτή θα έπρεπε να συνδυάζει κι άλλου είδους πληροφορίες, όπως για παράδειγμα μια «μετάφραση» της παρτιτούρας, ώστε να είναι προσιτή ακόμη και από μη ειδικούς στον τομέα, καλύπτοντας, έτσι, και εκπαιδευτικούς σκοπούς.

Κλείνοντας, θα μπορούσαμε να πούμε ότι η παρούσα διπλωματική εργασία, είναι απλώς ένας σπόρος, από όπου με την κατάλληλη «φροντίδα», μπορεί να ανθίσει ένα όμορφο δέντρο, με πολλά παρακλάδια.



## Βιβλιογραφία

---

- [1] C. Sachs, *World History of the Dance*, First Edition. New York: W. W. Norton & Company, 1963.
- [2] G. Fountzoulas, E. Nikolaki, M. Koutsouba, και V. Tyronola, ‘Η μορφολογική μέθοδος διδασκαλίας του ελληνικού παραδοσιακού χορού ως μέσο ανάπτυξης των δεξιοτήτων της κριτικής σκέψης’, τ. 1, σσ. 110–121, Ιουνίου 2021.
- [3] Β. Καρφής και Μ. Ζιάκα, *Ο Ελληνικός Παραδοσιακός Χορός στην Εκπαίδευση*. Θεσσαλονίκη: Διάπλους, 2009.
- [4] Μ. Κουτσούμπα, *ΣΗΜΕΙΟΓΡΑΦΙΑ ΤΗΣ ΧΟΡΕΥΤΙΚΗΣ ΚΙΝΗΣΗΣ: Το πέρασμα από την προϊστορία στην ιστορία του χορού*. Αθήνα: Εκδόσεις Προπομπός, 2005.
- [5] K. El Raheb και Y. Ioannidis, ‘Dance in the World of Data and Objects’, Ιανουαρίου 2013, σσ. 192–204. doi: 10.1007/978-3-642-40050-6\_17.
- [6] Πανδέκτης ©, ‘Πανδέκτης - Ο κόσμος του ελληνικού χορού’, Πανδέκτης. Ημερομηνία πρόσβασης: 19 Σεπτέμβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <https://dance-pandect.gr/>
- [7] ‘Ερευνητικό Πρόγραμμα 'Θράκη - Μακεδονία' - Εισαγωγή’. Ημερομηνία πρόσβασης: 19 Σεπτέμβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <http://epth.sfm.gr/>
- [8] H. Matsuyama, S. Aoki, T. Yonezawa, K. Hiroi, K. Kaji, και N. Kawaguchi, ‘Deep Learning for Ballroom Dance Recognition: A Temporal and Trajectory-Aware Classification Model With Three-Dimensional Pose Estimation and Wearable Sensing’, *IEEE Sensors Journal*, τ. 21, τχ. 22, σσ. 25437–25448, Αυγούστου 2021, doi: 10.1109/JSEN.2021.3098744.
- [9] S. Dewan, S. Agarwal, και N. Singh, ‘A deep learning pipeline for Indian dance style classification’, στο *Tenth International Conference on Machine Vision (ICMV 2017)*, SPIE, Απριλίου 2018, σσ. 265–273. doi: 10.1117/12.2309445.
- [10] J.-H. Kim, G.-S. Hong, B.-G. Kim, και D. P. Dogra, ‘deepGesture: Deep learning-based gesture recognition scheme using motion sensors’, *Displays*, τ. 55, σσ. 38–45, Δεκεμβρίου 2018, doi: 10.1016/j.displa.2018.08.001.
- [11] K. Cho και X. Chen, ‘Classifying and visualizing motion capture sequences using deep neural networks’, στο *2014 International Conference on Computer Vision Theory and Applications (VISAPP)*, Ιανουαρίου 2014, σσ. 122–130.

- [12] J. P. Ferreira κ.ά., ‘Learning to dance: A graph convolutional adversarial network to generate realistic dance motions from audio’, *Computers & Graphics*, τ. 94, σσ. 11–21, Φεβρουαρίου 2021, doi: 10.1016/j.cag.2020.09.009.
- [13] W. Li, B. Ren, H. Xu, S. Cao, και Y. Xie, ‘AutoDance: Music Driven Dance Generation’, στο *2021 International Symposium on Artificial Intelligence and its Application on Media (ISAIAM)*, Φεβρουαρίου 2021, σσ. 55–59. doi: 10.1109/ISAIAM53259.2021.00018.
- [14] ‘LabanWriter | Department of Dance’. Ημερομηνία πρόσβασης: 4 Οκτώβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <https://dance.osu.edu/research/dnb/labanner-writer>
- [15] ‘Labanatory’. Ημερομηνία πρόσβασης: 4 Οκτώβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <http://www.labanatory.com/>
- [16] A. Adamson, ‘Calaban’. Ημερομηνία πρόσβασης: 5 Οκτώβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <http://web.bham.ac.uk/calaban/frame.htm>
- [17] ‘LED & LINTEL: A Windows Editor/Interpreter for Labanotation’. Ημερομηνία πρόσβασης: 5 Οκτώβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <http://satd.com.au/don/pubs/led.html>
- [18] K. Kojima, K. Hachimura, και M. Nakamura, ‘LabanEditor: Graphical editor for dance notation’, Φεβρουαρίου 2002, σσ. 59–64. doi: 10.1109/ROMAN.2002.1045598.
- [19] L. Wilke, T. Calvert, R. Ryman, και I. Fox, ‘From dance notation to human animation: The LabanDancer project’, *Journal of Visualization and Computer Animation*, τ. 16, σσ. 201–211, Ιουλίου 2005, doi: 10.1002/cav.90.
- [20] M. Coyle, D. S. Maranan, και T. Calvert, ‘A Tool for Translating Dance Notation to Animation’, Ιανουαρίου 2002. doi: 10.13140/RG.2.1.4418.5762.
- [21] M. Nakamura, ‘AN XML REPRESENTATION OF LABANOTATION, LabanXML, AND ITS IMPLEMENTATION ON THE NOTATION EDITOR LabanEditor2’, 2006.
- [22] J. Hatol, ‘MovementXML: A representation of semantics of human movement based on Labanotation’, 2006. Ημερομηνία πρόσβασης: 5 Οκτώβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <https://www.semanticscholar.org/paper/MovementXML%3A-A-representation-of-semantics-of-human-Hatol/2be997c6db9cd5ba24f707ab64fcaa30d7bd96c6>
- [23] K. El Raheb και Y. Ioannidis, ‘A Labanotation Based Ontology for Representing Dance Movement’, στο *Gesture and Sign Language in Human-Computer Interaction and Embodied Communication*, E. Efthimiou, G. Kouroupetroglou, και S.-E. Fotinea, Επιμ., στο *Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer, 2012, σσ. 106–117. doi: 10.1007/978-3-642-34182-3\_10.

- [24] ‘The Method of Dance Movement Segmentation and Labanotation Generation Based on Rhythm | IEEE Journals & Magazine | IEEE Xplore’. Ημερομηνία πρόσβασης: 5 Οκτώβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <https://ieeexplore.ieee.org/document/9357349>
- [25] J. Wang, Z. Miao, H. Guo, Z. Zhou, και H. Wu, ‘Using automatic generation of Labanotation to protect folk dance’, *JEI*, τ. 26, τχ. 1, σ. 011028, Φεβρουαρίου 2017, doi: 10.1117/1.JEI.26.1.011028.
- [26] K. Ikeuchi, Z. Yan, Z. Ma, Y. Sato, M. Nakamura, και S. Kudoh, ‘Describing upper body motions based on the Labanotation for learning-from-observation robots’, arXiv.org. Ημερομηνία πρόσβασης: 5 Οκτώβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <https://arxiv.org/abs/1609.05429v1>
- [27] K. Hachimura και M. Nakamura, ‘Method of generating coded description of human body motion from motion-captured data’, στο *Proceedings 10th IEEE International Workshop on Robot and Human Interactive Communication. ROMAN 2001 (Cat. No.01TH8591)*, Σεπτεμβρίου 2001, σσ. 122–127. doi: 10.1109/ROMAN.2001.981889.
- [28] W. Choensawat, M. Nakamura, και K. Hachimura, ‘GenLaban: A tool for generating Labanotation from motion capture data’, *Multimedia Tools and Applications*, τ. 74, Ιανουαρίου 2014, doi: 10.1007/s11042-014-2209-6.
- [29] Z. Zhou, Z. Miao, και J. Wang, ‘A system for automatic generation of labanotation from motion capture data’, στο *2016 IEEE 13th International Conference on Signal Processing (ICSP)*, Αυγούστου 2016, σσ. 1031–1034. doi: 10.1109/ICSP.2016.7877986.
- [30] M. Li, Z. Miao, και C. Ma, ‘Automatic Labanotation Generation from Motion-Captured Data Based on Hidden Markov Models’, στο *2017 4th IAPR Asian Conference on Pattern Recognition (ACPR)*, Αυγούστου 2017, σσ. 793–798. doi: 10.1109/ACPR.2017.55.
- [31] S. Hao, Z. Miao, J. Wang, W. Xu, και Q. Zhang, ‘Labanotation Generation Based on Bidirectional Gated Recurrent Units with Joint and Line Features’, στο *2019 IEEE International Conference on Image Processing (ICIP)*, Σεπτεμβρίου 2019, σσ. 4265–4269. doi: 10.1109/ICIP.2019.8803659.
- [32] X. Zhang, Z. Miao, X. Yang, και Q. Zhang, ‘An Efficient Method for Automatic Generation of Labanotation Based on Bi-Directional LSTM’, *J. Phys.: Conf. Ser.*, τ. 1229, τχ. 1, σ. 012031, Φεβρουαρίου 2019, doi: 10.1088/1742-6596/1229/1/012031.
- [33] *Μουσική και Σώμα*. Αθήνα: Νάσος, 2022.
- [34] Ε. Κουρουπή, *Χορός, Σώμα, Κίνηση*. Αθήνα: Νεφέλη, 1999.
- [35] Π. Ηλιάδης, *Εγχειρίδιο Σύγχρονου Χορού*. Αθήνα, 2020.

- [36] ‘Grand Finale’, Hofesh Shechter Company. Ημερομηνία πρόσβασης: 8 Σεπτέμβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <https://hofesh.co.uk/productions/repertoire/grand-finale/>
- [37] Β. Τυροβολά, *Ο Ελληνικός Χορός: Μια Διαφορετική Προσέγγιση*. Αθήνα: Εκδόσεις Gutenberg, 2001.
- [38] Μ. Κουτσούμπα, ‘Η διδασκαλία του ελληνικού παραδοσιακού χορού σε σύγχρονα εκπαιδευτικά πλαίσια’, στο *Ελληνικός παραδοσιακός χορός. Θεωρήσεις για το λόγο, τη γραφή και τη διδασκαλία του*, Αθήνα: Εκδόσεις Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών, 2010, σσ. 101–126.
- [39] Ι. Θεοδοσίου, ‘Ο Ελληνικός Παραδοσιακός Χορός στην Πρωτοβάθμια Εκπαίδευση: Μια Πρόταση Διδασκαλίας’, Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών, Αθήνα, 2017.
- [40] ‘Μουσική – Χοροί της Θράκης – Κηπιώτες Έβρου’. Ημερομηνία πρόσβασης: 8 Σεπτέμβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <https://kipiotesevrou.com/mousikes-thrakis/>
- [41] Μ. Κουτσούμπα, ‘Η διδασκαλία του ελληνικού λαϊκού παραδοσιακού χορού στα σύγχρονα εκπαιδευτικά πλαίσια.’, παρουσιάστηκε στο Λαϊκός Πολιτισμός και Εκπαίδευση, Πρακτικά 1ου Διεθνούς Εκπαιδευτικού Συνεδρίου, 2007, σσ. 1–9.
- [42] Γ. Λυκεσάς, ‘Η διδασκαλία των ελληνικών παραδοσιακών χορών στη πρωτοβάθμια εκπαίδευση με τη μέθοδο της μουσικοκινητικής αγωγής’, Διδακτορική Διατριβή, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης (ΑΠΘ). Τμήμα Επιστήμης Φυσικής Αγωγής και Αθλητισμού. Τομέας Ανθρωπιστικών Σπουδών, 2002. Ημερομηνία πρόσβασης: 7 Σεπτέμβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <http://hdl.handle.net/10442/hedi/31490>
- [43] Β. Τυροβολά, ‘Η έννοια του αυτοσχεδιασμού στην ελληνική λαϊκή δημιουργία’, στο *Παραδοσιακός χορός και λαϊκή δημιουργία*, Κ. Σαχινίδης (Επιμ.), Αθήνα: Παπαζήσης, 1999, σσ. 99–132.
- [44] Ι. Loutzaki, ‘Ανθρωπολογία του Χορού και Σημειογραφία. Η Κινησιογραφία του Λάμπαν. Μια αναφορά στη χρήση της στην Ελλάδα’, τ. 93, σ. σελ. 32, Δεκεμβρίου 2004.
- [45] A. H. Guest, *Dance notation: the process of recording movement on paper*, 1. publ. London: Dance Books, 1984.
- [46] N. Xie, Z. Miao, και J. Wang, ‘Skeleton-Based Labanotation Generation Using Multi-model Aggregation’, στο *Pattern Recognition*, S. Palaiiahnakote, G. Sanniti di Baja, L. Wang, και W. Q. Yan, Επιμ., στο *Lecture Notes in Computer Science*. Cham: Springer International Publishing, 2020, σσ. 554–565. doi: 10.1007/978-3-030-41299-9\_43.



- [47] F. Özen, D. Tukul, και G. Dimirovski, ‘Synchronized dancing of an industrial manipulator and humans with arbitrary music’, *Acta Polytechnica Hungarica*, τ. 14, σσ. 151–169, Ιανουαρίου 2017, doi: 10.12700/APH.14.2.2017.2.8.
- [48] J. Birringer, ‘Dance and Media Technologies’, *PAJ: A Journal of Performance and Art*, τ. 24, σσ. 84–93, Ιανουαρίου 2002, doi: 10.1162/152028101753401811.
- [49] Κ. Καμπανέλλη, ‘Η χρήση κοστούμιών και σκηνικών σε παραστάσεις χορού’, στο *Η τέχνη του χορού σήμερα: εκπαίδευση, παραγωγή, παράσταση: πρακτικά συνεδρίου*, 2002, σσ. 91–96.
- [50] R. Burt, *Judson Dance Theater: performative traces*. London New York: Routledge, 2006.
- [51] ‘TBDC :: Homemade (1966)’. Ημερομηνία πρόσβασης: 13 Σεπτέμβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <https://trishabrowncompany.org/repertory/homemade.html>
- [52] Μ. Πανδρακλάκη, ‘Η συμβολή της τεχνολογίας στο σύγχρονο χορό από τις αρχές του 20ου αιώνα μέχρι σήμερα.’, Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών, Αθήνα, 2013.
- [53] M. R. Andrade, ‘Merce Cunningham and his use of Technology for Dance’, *Academia*, σσ. 3–4, 2018.
- [54] Α. Παπαδοπούλου, ‘Διαδραστική Πολυμεσική Εφαρμογή με θέμα 'Χορός και νέες Τεχνολογίες'.’, 2010.
- [55] S. Fdili Alaoui, ‘Dance Gesture Analysis and Visual Feedback based on Physical Models : Contributions of Movement Qualities in Whole Body Interaction’, Δεκεμβρίου 2012.
- [56] O. Grillmeyer, ‘Artificial Intelligence’, στο *Exploring Computer Science with Scheme*, O. Grillmeyer, Επιμ., στο Undergraduate Texts in Computer Science. , New York, NY: Springer, 1998, σσ. 411–474. doi: 10.1007/978-1-4757-2937-5\_15.
- [57] S. Russell και P. Norvig, *Τεχνητή Νοημοσύνη - Μια σύγχρονη προσέγγιση*. Αθήνα: Κλειδάριθμος, 2005.
- [58] A. Kaplan και M. Haenlein, ‘Siri, Siri, in my hand: Who’s the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence’, *Business Horizons*, τ. 62, τχ. 1, σσ. 15–25, Ιανουαρίου 2019, doi: 10.1016/j.bushor.2018.08.004.
- [59] M. Haenlein και A. Kaplan, ‘A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence’, *California Management Review*, τ. 61, τχ. 4, σσ. 5–14, Αυγούστου 2019, doi: 10.1177/0008125619864925.
- [60] M. Mohri, A. Rostamizadeh, και A. Talwalkar, *Foundations of machine learning*, Second edition. στο Adaptive computation and machine learning. Cambridge, Massachusetts London, England: The MIT Press, 2018.

- [61] Α. Γεωργούλη, ‘Μηχανική Μάθηση’, στο *Τεχνητή Νοημοσύνη [Προπτυχιακό εγχειρίδιο]*, Κάλλιπος, Ανοικτές Ακαδημαϊκές Εκδόσεις, 2016. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <http://hdl.handle.net/11419/3382>
- [62] A. L. Samuel, ‘Some studies in machine learning using the game of checkers’, *IBM Journal of Research and Development*, τ. 44, τχ. 1.2, σσ. 206–226, Ιανουαρίου 2000, doi: 10.1147/rd.441.0206.
- [63] T. M. Mitchell, *Machine Learning*. στο McGraw-Hill series in computer science. New York: McGraw-Hill, 1997.
- [64] G. Goel, ‘Human Learning vs Machine Learning’, Medium. Ημερομηνία πρόσβασης: 2 Σεπτέμβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <https://towardsdatascience.com/human-learning-vs-machine-learning-dfa8fe421560>
- [65] Β. Βερύκιος, Β. Καγκλής, και Η. Σταυρόπουλος, ‘Συσταδοποίηση’, στο *Η επιστήμη των δεδομένων μέσα από τη γλώσσα R [Προπτυχιακό εγχειρίδιο]*, Κάλλιπος, Ανοικτές Ακαδημαϊκές Εκδόσεις, 2016. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <http://hdl.handle.net/11419/2972>
- [66] K. J. Cios, R. W. Swiniarski, W. Pedrycz, και L. A. Kurgan, ‘Unsupervised Learning: Association Rules’, στο *Data Mining: A Knowledge Discovery Approach*, K. J. Cios, R. W. Swiniarski, W. Pedrycz, και L. A. Kurgan, Επιμ., Boston, MA: Springer US, 2007, σσ. 289–306. doi: 10.1007/978-0-387-36795-8\_10.
- [67] Ι. Βλαχάβας, Π. Κεφάλας, Ν. Βασιλειάδης, Φ. Κόκκορας, και Η. Σακελλαρίου, *Τεχνητή Νοημοσύνη*. Θεσσαλονίκη: Εκδόσεις Πανεπιστημίου Μακεδονίας, 2020.
- [68] Y. LeCun, Y. Bengio, και G. Hinton, ‘Deep Learning’, *Nature*, τ. 521, σσ. 436–44, Μαΐου 2015, doi: 10.1038/nature14539.
- [69] D. Jordens, ‘What is Deep Learning?’, Medium. Ημερομηνία πρόσβασης: 2 Σεπτέμβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <https://towardsdatascience.com/what-is-deep-learning-33a32858ea6d>
- [70] Σ. Φωτόπουλος και Α. Μακεδόνας, ‘Τεχνικές Μείωσης Διαστάσεων [PowerPoint slides]’, παρουσιάστηκε στο ELLAB, Πανεπιστήμιο Πατρών, 2008. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: [http://www.hep.upatras.gr/class/download/eyf\\_an\\_ded\\_anagn\\_prot/dimensionality%20reduction.pdf](http://www.hep.upatras.gr/class/download/eyf_an_ded_anagn_prot/dimensionality%20reduction.pdf)
- [71] K. P. Murphy, *Machine learning: a probabilistic perspective*, 4. print. (fixed many typos). στο Adaptive computation and machine learning series. Cambridge, Mass.: MIT Press, 2013.
- [72] R. Duda, P. Hart, και D. G. Stork, ‘Pattern Classification’, στο *Wiley Interscience*, τ. xx, 2001, σ. 15.

- [73] S. Theodoridis και K. Koutroumbas, *ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ*. Εκδόσεις Π. Χ. Πασχαλίδης, 2012.
- [74] ‘Principal Component Analysis (PCA) Explained | Built In’. Ημερομηνία πρόσβασης: 4 Σεπτέμβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <https://builtin.com/data-science/step-step-explanation-principal-component-analysis>
- [75] S. Haykin, *Νευρωνικά Δίκτυα & Μηχανική Μάθηση*, 3ο έκδ. Αθήνα: Εκδόσεις Παπασωτηρίου, 2010.
- [76] Α. Γ. Βραχάτης, ‘Ανάλυση Δεδομένων [PowerPoint slides]’, Πανεπιστήμιο Θεσσαλίας, 2018. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <http://archive.eclass.uth.gr/eclass/modules/document/file.php/DIB263/ΔΙΑΛΕΞΕΙΣ/PR%20-%202.pdf>
- [77] B. S. Everitt, S. Landau, M. Leese, και D. Stahl, *Cluster analysis*, 5th ed. στο Wiley series in probability and statistics, no. 848. Chichester: Wiley, 2011.
- [78] P. Hansen και B. Jaumard, ‘Cluster analysis and mathematical programming’, *Mathematical Programming*, τ. 79, τχ. 1, σσ. 191–215, Οκτωβρίου 1997, doi: 10.1007/BF02614317.
- [79] S. Kaushik, ‘Clustering | Introduction, Different Methods, and Applications (Updated 2023)’, Analytics Vidhya. Ημερομηνία πρόσβασης: 2 Σεπτέμβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <https://www.analyticsvidhya.com/blog/2016/11/an-introduction-to-clustering-and-different-methods-of-clustering/>
- [80] M. Kashyap, S. Gogoi, R. Prasad, και Student, *A Comparative Study on Partition-based Clustering Methods*. 2018.
- [81] ‘Clustering Algorithms | Machine Learning’, Google for Developers. Ημερομηνία πρόσβασης: 2 Σεπτέμβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <https://developers.google.com/machine-learning/clustering/clustering-algorithms>
- [82] A. I. Moreno, ‘Evaluating socio-economical indicators with exploratory analysis and hierarchical clustering’, Medium. Ημερομηνία πρόσβασης: 2 Σεπτέμβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <https://levelup.gitconnected.com/evaluating-socio-economical-indicators-with-exploratory-analysis-and-hierarchical-clustering-af1575782141>
- [83] S. Bandaru και K. Deb, ‘Towards automating the discovery of certain innovative design principles through a clustering-based optimization technique’, *Engineering Optimization*, τ. 43, σσ. 911–941, Σεπτεμβρίου 2011, doi: 10.1080/0305215X.2010.528410.
- [84] A. Banerjee και H. Shan, ‘Model-Based Clustering’, στο *Encyclopedia of Machine Learning*, C. Sammut και G. I. Webb, Επιμ., Boston, MA: Springer US, 2010, σσ. 686–689. doi: 10.1007/978-0-387-30164-8\_554.

- [85] D. Xu και Y. Tian, ‘A Comprehensive Survey of Clustering Algorithms’, *Ann. Data. Sci.*, τ. 2, τχ. 2, σσ. 165–193, Ιουνίου 2015, doi: 10.1007/s40745-015-0040-1.
- [86] B. B. & B. Greenwell, *Chapter 22 Model-based Clustering | Hands-On Machine Learning with R*. Ημερομηνία πρόσβασης: 2 Σεπτέμβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <https://bradleyboehmke.github.io/HOML/model-clustering.html>
- [87] ‘K-Mean clustering and its real use cases...’ Ημερομηνία πρόσβασης: 2 Σεπτέμβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <https://www.linkedin.com/pulse/k-mean-clustering-its-real-use-cases-yash-labhsetwar>
- [88] ‘2.3. Clustering’, scikit-learn. Ημερομηνία πρόσβασης: 2 Σεπτέμβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <https://scikit-learn/stable/modules/clustering.html>
- [89] R. Xu και D. Wunsch, ‘Survey of Clustering Algorithms’, *Neural Networks, IEEE Transactions on*, τ. 16, σσ. 645–678, Ιουνίου 2005, doi: 10.1109/TNN.2005.845141.
- [90] A. Tarekegn, K. Michalak, και M. Giacobini, ‘Cross-Validation Approach to Evaluate Clustering Algorithms: An Experimental Study Using Multi-Label Datasets’, *SN Computer Science*, τ. 1, Αυγούστου 2020, doi: 10.1007/s42979-020-00283-z.
- [91] E. Rendón, I. Abundez, A. Arizmendi, και E. M. Quiroz, ‘Internal versus external cluster validation indexes’, *International Journal of Computers and Communications*, τ. 5, σσ. 27–34, Ιανουαρίου 2011.
- [92] B. Mehlig, *Machine learning with neural networks*. 2021. doi: 10.1017/9781108860604.
- [93] Π. Αργυράκης, *Νευρωνικά Δίκτυα και Εφαρμογές*, τ. Β’. Πάτρα: Ελληνικό Ανοικτό Πανεπιστήμιο, 2001.
- [94] N. Negnevitsky, *Τεχνητή Νοημοσύνη*. Θεσσαλονίκη: Εκδόσεις Τζιόλα, 2018.
- [95] Rumelhart, D. E., J. McClelland, και J. L., *Parallel distributed processing: explorations in the microstructure of cognition. Volume 1. Foundations*. 1986.
- [96] W. S. McCulloch και W. Pitts, ‘A logical calculus of the ideas immanent in nervous activity’, *Bulletin of Mathematical Biophysics*, τ. 5, τχ. 4, σσ. 115–133, Δεκεμβρίου 1943, doi: 10.1007/BF02478259.
- [97] F. Rosenblatt, ‘The perceptron: a probabilistic model for information storage and organization in the brain’, *Psychol Rev*, τ. 65, τχ. 6, σσ. 386–408, Νοεμβρίου 1958, doi: 10.1037/h0042519.
- [98] J. J. Hopfield, ‘Neural networks and physical systems with emergent collective computational abilities.’, *Proc Natl Acad Sci U S A*, τ. 79, τχ. 8, σσ. 2554–2558, Απριλίου 1982.
- [99] M. Y. Kiang, ‘Neural Networks’, στο *Encyclopedia of Information Systems*, H. Bidgoli, Επιμ., New York: Elsevier, 2003, σσ. 303–315. doi: 10.1016/B0-12-227240-4/00121-0.

- [100] E. Pekel και S. Kara, ‘A COMPREHENSIVE REVIEW FOR ARTIFICIAL NEURAL NETWORK APPLICATION TO PUBLIC TRANSPORTATION’, *Sigma Journal of Engineering and Natural Sciences*, τ. 35, σσ. 157–179, Μαρτίου 2017.
- [101] ‘Single Layer Perceptron in TensorFlow - Javatpoint’. Ημερομηνία πρόσβασης: 2 Σεπτέμβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <https://www.javatpoint.com/single-layer-perceptron-in-tensorflow>
- [102] B. Mostafa, N. El-Attar, S. Abd-Elhafeez, και W. Awad, ‘Machine and Deep Learning Approaches in Genome: Review Article’, *Alfarama Journal of Basic & Applied Sciences*, Αυγούστου 2020, doi: 10.21608/ajbas.2020.34160.1023.
- [103] C. S. Nageswari, M. N. V. Kumar, C. Raveena, J. S. Sharma, και M. Y. Devi, ‘An Identification and Classification of Thyroid Diseases Using Deep Learning Methodology’, *Revista Gestão Inovação e Tecnologias*, τ. 11, τχ. 2, σ. 2004, Ιανουαρίου 2021.
- [104] I. Goodfellow, Y. Bengio, και A. Courville, *Deep learning*. στο Adaptive computation and machine learning. Cambridge, Massachusetts: The MIT Press, 2016.
- [105] I. Tabian, H. Fu, και Z. Sharif Khodaei, ‘A Convolutional Neural Network for Impact Detection and Characterization of Complex Composite Structures’, *Sensors*, τ. 19, σ. 4933, Νοεμβρίου 2019, doi: 10.3390/s19224933.
- [106] S. Chetlur κ.ά., ‘cuDNN: Efficient Primitives for Deep Learning’, Οκτωβρίου 2014.
- [107] D. H. Hubel και T. N. Wiesel, ‘Receptive fields of single neurones in the cat’s striate cortex’, *J Physiol*, τ. 148, τχ. 3, σσ. 574–591, Οκτωβρίου 1959.
- [108] D. H. Hubel και T. N. Wiesel, ‘Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex’, *J Physiol*, τ. 160, τχ. 1, σσ. 106-154.2, Ιανουαρίου 1962.
- [109] D. H. Hubel και T. N. Wiesel, ‘Receptive fields and functional architecture of monkey striate cortex’, *J Physiol*, τ. 195, τχ. 1, σσ. 215–243, Μαρτίου 1968, doi: 10.1113/jphysiol.1968.sp008455.
- [110] K. Fukushima, ‘Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position’, *Biol Cybern*, τ. 36, τχ. 4, σσ. 193–202, 1980, doi: 10.1007/BF00344251.
- [111] Y. Lecun, L. Bottou, Y. Bengio, και P. Haffner, ‘Gradient-Based Learning Applied to Document Recognition’, *Proceedings of the IEEE*, τ. 86, σσ. 2278–2324, Δεκεμβρίου 1998, doi: 10.1109/5.726791.
- [112] A. Krizhevsky, I. Sutskever, και G. Hinton, ‘ImageNet Classification with Deep Convolutional Neural Networks’, *Neural Information Processing Systems*, τ. 25, Ιανουαρίου 2012, doi: 10.1145/3065386.

- [113] S. Saha, ‘A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way’, Medium. Ημερομηνία πρόσβασης: 4 Σεπτέμβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>
- [114] R. Hachilif, R. Baghdadi, και F. Benhamida, ‘Graduation Thesis Implementing and Optimizing Neural Networks using Tiramisu’, 2019.
- [115] A. Dertat, ‘Applied Deep Learning - Part 4: Convolutional Neural Networks’, Medium. Ημερομηνία πρόσβασης: 4 Σεπτέμβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2>
- [116] H. Gholamalinezhad και H. Khosravi, ‘Pooling Methods in Deep Neural Networks, a Review’, 2020, doi: 10.48550/ARXIV.2009.07485.
- [117] B. Cunha, C. Droz, A. Zine, S. Foulard, και M. Ichchou, ‘A Review of Machine Learning Methods Applied to Structural Dynamics and Vibroacoustic’, *Mechanical Systems and Signal Processing*, τ. 200, σ. 110535, Οκτωβρίου 2023, doi: 10.1016/j.ymssp.2023.110535.
- [118] S. J. Pan και Q. Yang, ‘A Survey on Transfer Learning’, *IEEE Transactions on Knowledge and Data Engineering*, τ. 22, τχ. 10, σσ. 1345–1359, Ιουλίου 2010, doi: 10.1109/TKDE.2009.191.
- [119] L. Torrey και J. Shavlik, ‘Transfer Learning’., στο *Handbook of Research on Machine Learning Applications and Trends*, E. S. Olivas, J. D. M. Guerrero, M. Martinez-Sober, J. R. Magdalena-Benedito, και A. J. Serrano López, Επιμ., IGI Global, 2010, σσ. 242–264. doi: 10.4018/978-1-60566-766-9.ch011.
- [120] ‘Keras: Deep Learning for humans’. Ημερομηνία πρόσβασης: 11 Οκτώβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <https://keras.io/>
- [121] ‘TensorFlow’. Ημερομηνία πρόσβασης: 13 Οκτώβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <https://www.tensorflow.org/>
- [122] F. Pedregosa κ.ά., ‘Scikit-learn: Machine Learning in Python’, *Journal of Machine Learning Research*, τ. 12, τχ. 85, σσ. 2825–2830, 2011.
- [123] ‘NumPy’. Ημερομηνία πρόσβασης: 14 Οκτώβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <https://numpy.org/>
- [124] ‘pandas - Python Data Analysis Library’. Ημερομηνία πρόσβασης: 14 Οκτώβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <https://pandas.pydata.org/>
- [125] ‘opencv-python: Wrapper package for OpenCV python bindings.’ Ημερομηνία πρόσβασης: 14 Οκτώβριος 2023. [MacOS, Microsoft :: Windows, POSIX, Unix]. Διαθέσιμο στο: <https://github.com/opencv/opencv-python>

- [126] ‘Matplotlib — Visualization with Python’. Ημερομηνία πρόσβασης: 14 Οκτώβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <https://matplotlib.org/>
- [127] K. Simonyan και A. Zisserman, ‘Very Deep Convolutional Networks for Large-Scale Image Recognition’, 2014, doi: 10.48550/ARXIV.1409.1556.
- [128] J. Yoon κ.ά., ‘Classifying central serous chorioretinopathy subtypes with a deep neural network using optical coherence tomography images: a cross-sectional study’, *Sci Rep*, τ. 12, σ. 422, Ιανουαρίου 2022, doi: 10.1038/s41598-021-04424-z.
- [129] K. He, X. Zhang, S. Ren, και J. Sun, ‘Deep Residual Learning for Image Recognition’, 2015, doi: 10.48550/ARXIV.1512.03385.
- [130] C. Szegedy κ.ά., ‘Going Deeper with Convolutions’, 2014, doi: 10.48550/ARXIV.1409.4842.
- [131] F. Chollet, ‘Xception: Deep Learning with Depthwise Separable Convolutions’, στο *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI: IEEE, Ιουλίου 2017, σσ. 1800–1807. doi: 10.1109/CVPR.2017.195.
- [132] S.-H. Tsang, ‘Review: Xception — With Depthwise Separable Convolution, Better Than Inception-v3 (Image...’, Medium. Ημερομηνία πρόσβασης: 2 Σεπτέμβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <https://towardsdatascience.com/review-xception-with-depthwise-separable-convolution-better-than-inception-v3-image-dc967dd42568>
- [133] N. N. Kachouie και M. Shutaywi, ‘Weighted Mutual Information for Aggregated Kernel Clustering’, *Entropy (Basel)*, τ. 22, τχ. 3, σ. 351, Μαρτίου 2020, doi: 10.3390/e22030351.
- [134] ‘Evaluation of clustering’. Ημερομηνία πρόσβασης: 2 Σεπτέμβριος 2023. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: <https://nlp.stanford.edu/IR-book/html/htmledition/evaluation-of-clustering-1.html>