



# ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ  
ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Βαθιά Νευρωνικά Δίκτυα για την Ανάλυση Ακτινογραφιών και  
Αξονικών Τομογραφιών Ασθενών με COVID-19

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Νικόλαος Η. Μπούας

Επιβλέπων : Στέφανος Κόλλιας

Καθηγητής Ε.Μ.Π

Αθήνα, Ιούλιος 2020





ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ  
ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Βαθιά Νευρωνικά Δίκτυα για την Ανάλυση Ακτινογραφιών και  
Αξονικών Τομογραφιών Ασθενών με COVID-19

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Νικόλαος Η. Μπούας

Επιβλέπων : Στέφανος Κόλλιας

Καθηγητής Ε.Μ.Π

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 24η Ιουλίου 2020.

.....  
Στέφανος Κόλλιας

Καθηγητής Ε.Μ.Π

.....  
Ανδρέας-Γεώργιος Σταφυλοπάτης

Καθηγητής Ε.Μ.Π

.....  
Γιώργος Στάμου

Αναπληρωτής Καθηγητής Ε.Μ.Π

Αθήνα, Ιούλιος 2020

.....

Νικόλαος Η. Μπούας

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

Copyright©Νικόλαος Η. Μπούας, 2020. Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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





## Περίληψη

Ο ραγδαία αυξανόμενος αριθμός κρουσμάτων COVID-19 και οι επιπτώσεις που έχει το γεγονός αυτό σε κοινωνικό επίπεδο, επισημάνουν την ανάγκη ανάπτυξης μηχανισμών για την γρήγορη και αποτελεσματική ανίχνευση των κρουσμάτων αυτών. Ο παραδοσιακός τρόπος ανίχνευσης είναι συχνά χρονοβόρος και ακριβός.

Στο πλαίσιο της διπλωματικής αυτής μελετάμε και αναπτύσσουμε έξυπνα συστήματα για την επιτυχή ανίχνευση κρουσμάτων με τη χρήση ακτινογραφιών ή αξονικών τομογραφιών.

Ένα βασικό μειονέκτημα των συστημάτων αυτών είναι ότι συχνά λειτουργούν ως "μαύρο κουτί", δηλαδή δε γνωρίζουμε γιατί λαμβάνουν την απόφαση που έλαβαν. Το πρόβλημα αυτό είναι σημαντικό να αντιμετωπιστεί, ιδιαίτερα όταν έχουμε να κάνουμε με συστήματα ιατρικής υποβοήθησης όπως αυτά που θα αναπτύξουμε. Για να το αντιμετωπίσουμε, εφαρμόσαμε μεθόδους ομαδοποίησης οι οποίες πέρα από την κατηγοριοποίηση των εικόνων, μπορούν εύκολα να παρέχουν πληροφορία που να εξηγεί τους λόγους για τους οποίους το σύστημα πήρε μία συγκεκριμένη απόφαση.

Τέλος, αναπτύσσουμε συστήματα ικανά να ταξινομήσουν εικόνες που προέρχονται από σύνολα δεδομένων που περιέχουν ακτινογραφίες αλλά και αξονικές τομογραφίες με κλασσικές μεθόδους μηχανικής μάθησης ή χρησιμοποιώντας ομαδοποίηση.

**Λέξεις κλειδιά:** Convolutional Neural Networks (CNNs) , Clustering, Covid-19, Ακτινογραφία, Αξονική Τομογραφία (CT)

## Abstract

The rapidly increasing number of COVID-19 cases and the social impacts of this phenomena, underline the importance of the development of mechanisms for quick and effective testing. Traditional testing methods are often time consuming and expensive.

In this diploma thesis, we study and develop intelligent systems capable of detecting COVID-19 cases from chest X-ray or CT images.

An important disadvantage these kind of systems usually have, is that they often operate as a "black box", namely we do not know why they take the decisions they take. It is important to tackle this problem, especially when our task concerns systems that aid clinicians in decision-making, like in our case. In order to tackle it, we use clustering methods, which can provide, aside from image classification, insights that explain why a system made a certain decision.

Lastly, we develop systems capable of classifying images that come from datasets that contain both X-rays and CTs, using traditional machine learning methods as well as clustering.

**Keywords:** Convolutional Neural Networks (CNNs) , Clustering, Covid-19, X-ray, Computed Tomography (CT)



## Περιεχόμενα

1.1 Εισαγωγή.....	11
1.2 Δομή Διπλωματικής .....	13
2. Θεωρητικό Υπόβαθρο .....	14
2.1 Νευρωνικά Δίκτυα .....	14
2.2 Συνελικτικά Νευρωνικά Δίκτυα .....	19
2.3 ResNet-50.....	22
2.4 Μεταφορά Μάθησης (Transfer Learning) .....	23
2.5 Συσταδοποίηση (Clustering) .....	25
3. Σύνολο Δεδομένων με Ακτινογραφίες.....	28
3.1 Covid-Next.....	28
3.2 Chest X-Ray Dataset .....	28
3.3 Χρήση Επιπλέον Επιπέδου (With vs Without Extra layer).....	30
3.4 Επανεκπαίδευση ή μη της Βασικής Δομής (Frozen vs Not Frozen Backbone) .....	31
3.5 Μετρικές ανά κλάση .....	32
3.6 Συσταδοποίηση σε 3 Κλάσεις .....	34
3.7 Κατηγοριοποίηση σε 2 Κλάσεις .....	37
3.8 Συσταδοποίηση σε 2 Κλάσεις .....	38
4. Σύνολο Δεδομένων με Αξονικές Τομογραφίες .....	40
4.1 Αρχιτεκτονική.....	40
4.2 COVID-CT Dataset .....	40
4.3 Χρήση Προ-εκπαιδευμένων ή μη Δικτύων (Pretrained vs Not-Pretrained) .....	42
4.4 Συσταδοποίηση στην Ανάλυση Αξονικών Τομογραφιών .....	44
5. Σύνολο Δεδομένων με Ακτινογραφίες και Αξονικές Τομογραφίες .....	46
5.1 Αρχιτεκτονική.....	46
5.2 Dataset .....	46
5.3 Χρήση Μόνο Ακτινογραφιών ή Ακτινογραφιών & Αξονικών Τομογραφιών.....	46
5.4 Σύγκριση της Χρήσης μόνο Αξονικών Τομογραφιών ή και Ακτινογραφιών.....	47
5.5 Ισορροπημένες Κατηγορίες Δεδομένων.....	49
5.6 Ισορροπημένα Είδη Δεδομένων .....	50
5.7 Ισορροπημένα Είδη και Κατηγορίες Δεδομένων.....	51
5.8 Ευθυγράμμιση του Είδους των Δεδομένων .....	53
5.9 Ευθυγράμμιση των Κατηγοριών των Δεδομένων .....	54

5.10 Ευθυγράμμιση των Κατηγοριών και του Είδους των Δεδομένων.....	56
5.11 Συσταδοποίηση.....	57
5.12 Συμπεράσματα.....	59
6. Πειράματα σε ελληνικά δεδομένα .....	61
6.1 Αρχιτεκτονική.....	61
6.2 Σύνολο Δεδομένων .....	61
6.3 Χωρίς επανεκπαίδευση .....	62
6.3.1 Με τεχνική "At least one" .....	62
6.3.2 Με τεχνική "Majority voting".....	63
6.4 Με επανεκπαίδευση.....	63
6.4.1 Με τεχνική "At least one" .....	64
6.4.2 Με τεχνική "Majority voting".....	64
6.5 Συσταδοποίηση.....	65
7. Σύνοψη.....	67
Βιβλιογραφία .....	68

## 1.1 Εισαγωγή

Η πανδημία της νόσου του κορονοϊού είναι μία τρέχουσα πανδημία που προκλήθηκε από τον κορονοϊό SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2). Η εξάπλωση ξεκίνησε τον Δεκέμβριο του 2019 από την πόλη Ουχάν, της Κίνας. Ο παγκόσμιος οργανισμός υγείας ανακήρυξε τον COVID-19 πανδημία στις 11 Μαρτίου του 2020. Ως τις 7 Ιουλίου το 2020, έχουν επιβεβαιωθεί 11,4 εκατομμύρια κρούσματα, σε περισσότερες από 188 χώρες και περιοχές ενώ έχουν παρατηρηθεί περισσότεροι από 534 χιλιάδες θάνατοι λόγω της νόσου.

Από τα παραπάνω δεδομένα γίνεται εύκολα εμφανής η ανάγκη για τη μείωση της ταχύτητας εξάπλωσης του νέου κορονοϊού. Ένας τρόπος για να το πετύχουμε αυτό είναι η έγκαιρη διάγνωση και απομόνωση των καινούριων κρουσμάτων. Προς το παρόν, εφαρμόζονται κατά κύριο λόγο δύο μέθοδοι διάγνωσης, συγκεκριμένα το τεστ αλυσιδωτής αντίδρασης πολυμεράσης και το τεστ αντισωμάτων.

Στο τεστ αλυσιδωτής αντίδρασης πολυμεράσης προσπαθούμε να ανιχνεύσουμε το RNA του ιού σε δείγματα που λαμβάνουμε συνήθως από την μύτη ή το στόμα. Με τον τρόπο αυτό μπορούμε να ανιχνεύσουμε αποτελεσματικά τρέχοντες μολύνσεις από κορονοϊό μέσα σε 6-8 ώρες. Το τεστ αυτό είναι αρκετά ακριβό και χρονοβόρο, ενώ απαιτεί και τον πολύτιμο χρόνο ιατρικού προσωπικού .

Το τεστ αντισωμάτων δουλεύει αρκετά διαφορετικά από το προηγούμενο αφού προσπαθεί να ανιχνεύσει την ανοσοποιητική αντίδραση του οργανισμού στον ιό, δηλαδή τα αντισώματα και όχι τον ίδιο τον ιό. Το τεστ αυτό είναι ιδιαίτερα χρήσιμο αν θέλουμε να προσδιορίσουμε τον βαθμό στον οποίο ένας πληθυσμός έχει αποκτήσει "ανοσία αγέλης" καθώς ανιχνεύει ασθενείς που έχουν αναρρώσει ή πάσχουν αυτή τη στιγμή από τη νόσο. Επειδή, όμως το τεστ δεν είναι προς το παρόν ικανό να διακρίνει ανάμεσα στις δύο αυτές περιπτώσεις καταλαβαίνουμε ότι έχει και αυτό σημαντικά μειονεκτήματα.

Στο πλαίσιο της διπλωματικής αυτής, μελετήσαμε έναν διαφορετικό τρόπο διάγνωσης, συγκεκριμένα, την ανίχνευση κρουσμάτων μέσω ακτινογραφιών στήθους και αξονικών τομογραφιών. Για το σκοπό αυτό μελετήσαμε και επεκτείναμε τις εργασίες των [1], [2], [3], [4], [5]. Χρησιμοποιήσαμε βαθειά συνελκτικά νευρωνικά δίκτυα (CNNs) ([6], [7], [8], [9], [10]). Αρχικά, δοκιμάσαμε τα μοντέλα σε σύνολα δεδομένων που περιείχαν ακτινογραφίες για την κατηγοριοποίηση σε 3 κλάσεις (Υγιείς, Πνευμονία και COVID-19). Έπειτα εφαρμόσαμε μεθόδους clustering πετυχαίνοντας έτσι τόσο την κατηγοριοποίηση όσο και την εξαγωγή πυρήνων που μπορούν να παρέχουν σημασιολογική πληροφορία που μας βοηθάει να καταλάβουμε γιατί κάποιο μοντέλο πήρε μία συγκεκριμένη απόφαση. Στη συνέχεια, εφαρμόσαμε παρόμοια διαδικασία αλλά αυτή τη φορά κατηγοριοποιώντας τα δεδομένα σε 2 κλάσεις ( Ασθενείς COVID-19 και μη-Ασθενείς COVID-19). Έπειτα, επαναλάβαμε τα παραπάνω πειράματα για σύνολα δεδομένων που περιείχαν αξονικές τομογραφίες. Αναπτύξαμε επίσης μοντέλα που είναι ικανά να κατηγοριοποιήσουν εικόνες προερχόμενες από τα δύο παραπάνω datasets ταυτόχρονα. Επειδή, υπήρχε μεγάλη διαφορά

ανάμεσα στον αριθμό δειγμάτων των δύο datasets αλλά και μέσα στα ίδια τα datasets , όσον αφορά τον αριθμό των δειγμάτων της κάθε κλάσης για να πετύχουμε το παραπάνω εφαρμόσαμε μεθόδους balancing και alignment που θα αναλύσουμε στη συνέχεια. Τέλος, χρησιμοποιώντας δεδομένα που προέρχονται από ελληνικά νοσοκομεία, επανεκπαίδευσαν τα μοντέλα ( fine-tuning) και τα δοκιμάσαμε στη διάγνωση ασθενών που πάσχουν από κορονοϊό μέσα από σειρές αξονικών τομογραφιών (CT series).

## 1.2 Δομή Διπλωματικής

Η διπλωματική αποτελείται από 6 κεφάλαια. Στο πρώτο κεφάλαιο κάνουμε μία σύντομη εισαγωγή όπου παρουσιάζουμε συνοπτικά το πρόβλημα με το οποίο ασχολείται η παρούσα διπλωματική και τις μεθοδολογίες που ακολουθήσαμε για να το αντιμετωπίσουμε. Στο δεύτερο κεφάλαιο παρουσιάζουμε κάποιες θεωρητικές έννοιες τις οποίες θα χρησιμοποιήσουμε στη συνέχεια για πληρέστερη κατανόηση της εργασίας. Στο τρίτο κεφάλαιο παρουσιάζουμε σύνολα δεδομένων, μοντέλα και πειράματα τα οποία αποσκοπούν στην κατηγοριοποίηση ακτινογραφιών ασθενών. Εφαρμόζουμε επίσης ομαδοποίηση (clustering) και συγκρίνουμε τις δύο τεχνικές. Στο τέταρτο κεφάλαιο, ακολουθούμε παρόμοια διαδικασία αυτή τη φορά όμως, για σύνολα δεδομένων που περιέχουν αξονικές τομογραφίες. Στο πέμπτο κεφάλαιο, εξετάζουμε τεχνικές που χρησιμοποιήσαμε για να αναπτύξουμε μοντέλα τα οποία είναι ικανά να ταξινομήσουν εικόνες που προέρχονται από σύνολα δεδομένων που περιέχουν εικόνες ακτινογραφιών και αξονικών τομογραφιών. Τέλος, στο έκτο κεφάλαιο, παρουσιάζουμε πειράματα τα οποία αφορούν δεδομένα που πήραμε από ελληνικά νοσοκομεία και περιέχουν σειρές αξονικών τομογραφιών ασθενών (CT series).

## 2. Θεωρητικό Υπόβαθρο

### 2.1 Νευρωνικά Δίκτυα

Αν και αλγόριθμοι που αφορούν τα νευρωνικά δίκτυα άρχισαν να αναπτύσσονται από το 1943, λόγω έλλειψης αρκετά ισχυρού υλικού (hardware) , οι αλγόριθμοι αυτοί δεν ήταν πρακτικά χρήσιμοι για αρκετές δεκαετίες. Την δεκαετία του 2000, με την ανάπτυξη του υλικού αυτού και την ανακάλυψη ακόμα πιο εξελιγμένων αλγορίθμων το ενδιαφέρον προς το πεδίο αυτό επανήλθε. Το 2009 ο Andrew Ng έδειξε ότι χρησιμοποιώντας GPU η εκπαίδευση των νευρωνικών δικτύων μπορεί να επιταχυνθεί κατά 100 φορές. Ακολούθησαν πολλές ανακαλύψεις στον τομέα , ενώ το 2012 ένα βαθύ νευρωνικό δίκτυο που αναπτύχθηκε από τον Krizhevsky και τους συνεργάτες του κέρδισε τον γνωστό διαγωνισμό Imagenet. Από πολλούς η νίκη αυτή θεωρείται ότι σηματοδότησε την επανάσταση της βαθιάς μάθησης ( [11] ).

Το δίκτυο που κέρδισε τον διαγωνισμό Imagenet το 2012, αλλά και τα νευρωνικά δίκτυα που χρησιμοποιούνται σήμερα αποτελούνται από πολλαπλά επίπεδα για να εξάγουν χαρακτηριστικά ολοένα και υψηλότερου επιπέδου από τα δεδομένα. Για παράδειγμα, στην επεξεργασία εικόνων, τα χαμηλά επίπεδα των δικτύων αυτών μπορεί να αναγνωρίζουν χαρακτηριστικά όπως είναι οι ακμές και οι γωνίες ενώ τα υψηλότερα επίπεδα εξάγουν χαρακτηριστικά πιο κοντά στην ανθρώπινη λογική , όπως είναι τα ψηφία , τα γράμματα ή τα πρόσωπα.

Τα προβλήματα που αντιμετωπίζουν τα νευρωνικά δίκτυα χωρίζονται σε δύο βασικές κατηγορίες:

- προβλήματα κατηγοριοποίησης (classification) και
- προβλήματα παλινδρόμησης (regression)

Τα προβλήματα κατηγοριοποίησης αφορούν την πρόβλεψη της ετικέτας μίας κλάσης για ένα δεδομένο παράδειγμα. Τέτοια προβλήματα μπορεί να είναι η αναγνώριση ενός μουσικού είδους από ένα κομμάτι μουσικής ή αναγνώριση του είδους ενός αντικειμένου σε μία εικόνα.

Τα προβλήματα παλινδρόμησης αφορούν την πρόβλεψη μίας τιμής για ένα παράδειγμα. Ένα τέτοιο πρόβλημα μπορεί να είναι η πρόβλεψη του κόστους ενός έργου δεδομένων κάποιων στοιχείων , όπως το είδος του έργου, το μέγεθός του, ο αριθμός ανθρώπων που χρειάζονται για την επίτευξη του κλπ.

Στην παρούσα διπλωματική θα ασχοληθούμε με προβλήματα ταξινόμησης που αφορούν ιατρικές εικόνες ( [12], [13], [14], [15]).

Τα συστήματα που χρησιμοποιούν νευρωνικά δίκτυα περιλαμβάνουν τυπικά δύο στάδια :

- Το στάδιο training
- Το στάδιο testing

Κατά την εκπαίδευση ενός νευρωνικού δικτύου χρειάζεται πληθώρα δεδομένων τα οποία δίνουμε στο νευρωνικό δίκτυο, και αυτό μέσω αλγορίθμων όπως είναι ο gradient descent ανανεώνει τις παραμέτρους προκειμένου να μάθει να αναγνωρίζει και να εξάγει χαρακτηριστικά χρήσιμα για το δεδομένο πρόβλημα.

Στο στάδιο του testing δοκιμάζουμε το νευρωνικό δίκτυο και μετράμε την απόδοσή του σε δεδομένα τα οποία δεν έχει "ξαναδεί". Η απόδοση μπορεί να μετρηθεί με διάφορες μετρικές και η επιλογή της κατάλληλης μετρικής εξαρτάται κάθε φορά από το πρόβλημα που έχουμε να αντιμετωπίσουμε. Οι βασικές μετρικές που χρησιμοποιούμε είναι οι :

- Accuracy
- Precision
- Recall
- F1 Score

Στη συνέχεια, κάνουμε μία σύντομη παρουσίαση των μετρικών αυτών.

Η μετρική accuracy αντιπροσωπεύει το ποσοστό των παραδειγμάτων τα οποία ταξινομήθηκαν σωστά. Με μία πρώτη ματιά η μετρική αυτή φαίνεται ότι αντιπροσωπεύει πολύ σωστά την απόδοση ενός μοντέλου. Ας εξετάσουμε όμως το παρακάτω παράδειγμα.

		Προβλέψεις	
		Αρνητικά	Θετικά
Πραγματικές Τιμές	Αρνητικά	998	0
	Θετικά	1	1

Εύκολα παρατηρούμε ότι η accuracy για το παραπάνω παράδειγμα είναι 99.9% . Χρησιμοποιώντας λοιπόν την μετρική αυτή θα έλεγε κανείς ότι το μοντέλο που έκανε τις παραπάνω προβλέψεις είναι καλό. Τι θα γινόταν όμως αν τα θετικά που αναφέρουμε παραπάνω αφορούσαν κρούσματα ενός ιού που μεταδίδεται ραγδαία ; Στην περίπτωση αυτή, αλλά και σε πολλές άλλες που μπορούμε να σκεφτούμε, το κόστος ενός θετικού

παραδείγματος που λανθασμένα έχει προβλεφθεί ως αρνητικό (false negative) είναι πολύ υψηλό. Καταλαβαίνουμε λοιπόν ότι η μετρική accuracy δεν αποτελεί πανάκεια. Αναζητώντας εναλλακτικές εισάγουμε στη συνέχεια δύο νέες μετρικές, το precision και το recall.

Οι τύποι για τον υπολογισμό των δύο αυτών μετρικών δίνονται παρακάτω:

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

Εικόνα 2.1: Υπολογισμός precision και recall (Πηγή: [16])

Αν εξετάσουμε λίγο περαιτέρω τον τύπο για το precision βλέπουμε ότι στην ουσία ο παρονομαστής αποτελεί το σύνολο των προβλεπόμενων θετικών. Πράγματι :

$$\begin{aligned} \text{Precision} &= \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \\ &= \frac{\text{True Positive}}{\text{Total Predicted Positive}} \end{aligned}$$

Εικόνα 2.2: Μετατροπή τύπου precision (Πηγή: [16])

Είναι μία καλή μετρική όταν το κόστος των false positive, δηλαδή των δειγμάτων που ταξινομούνται λανθασμένα ως θετικά, είναι υψηλό. Μία τέτοια περίπτωση μπορεί να είναι ένα σύστημα ανίχνευσης ανεπιθύμητων μηνυμάτων όπου ένα false positive θα σήμαινε ότι ένα ασφαλές email μήνυμα ανιχνεύθηκε ως ανεπιθύμητο και έτσι ο χρήστης να έχανε σημαντική πληροφορία που περιείχε το μήνυμα.

Αντίστοιχα, αν εξετάσουμε τον τύπο για το recall παρατηρούμε ότι αντιπροσωπεύει τον αριθμό των δειγμάτων που σωστά προβλέφθηκαν ως θετικά προς το σύνολο των πραγματικών θετικών :

$$\begin{aligned} \text{Recall} &= \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \\ &= \frac{\text{True Positive}}{\text{Total Actual Positive}} \end{aligned}$$

Εικόνα 2.3: Μετατροπή τύπου recall (Πηγή: [16])



Το recall αποτελεί μία καλή μετρική όταν το κόστος των false negatives , δηλαδή τα δείγματα που ταξινομούνται λανθασμένα ως αρνητικά είναι ακριβό. Ένα καλό παράδειγμα αποτελεί αυτό μίας μεταδοτικής ασθένειας, όπου μία θετική πρόβλεψη αντιστοιχεί σε έναν ασθενή που πάσχει από την ασθένεια και μία αρνητική σε έναν υγιή άνθρωπο. Στην περίπτωση αυτή, θα θέλαμε να αποφύγουμε οπωσδήποτε την περίπτωση ένας άνθρωπος που είναι άρρωστος να χαρακτηριστεί ως υγιείς αφού κάτι τέτοιο θα οδηγούσε σε γρήγορη εξάπλωση της ασθένειας.

Το F1 score είναι η τελευταία μετρική που θα εξετάσουμε και ο τύπος για τον υπολογισμό του είναι ο εξής :

$$F1 = 2 \times \frac{Precision * Recall}{Precision + Recall}$$

Εικόνα 2.4: Υπολογισμός F1 score (Πηγή: [16])

Το F1 score χρησιμοποιείται όταν θέλουμε να έχουμε μία ισορροπία ανάμεσα σε precision και recall. Μία εύλογη απορία θα ήταν τότε , ποια είναι η διαφορά του από το accuracy. Η διάφορα έγκειται στο γεγονός ότι το F1 score πέρα από την ισορροπία ανάμεσα στις δύο μετρικές , precision και recall, δίνει αντιπροσωπευτικά αποτελέσματα και όταν υπάρχουν μεγάλες διαφορές στην κατανομή των κλάσεων, έχουμε δηλαδή imbalanced datasets.

Επειδή στην περίπτωση μας έχουμε imbalanced datasets, ιδιαίτερα το σύνολο δεδομένων που περιέχει τις ακτινογραφίες, όπως θα δούμε στη συνέχεια , θα χρησιμοποιούμε κυρίως το F1 score ως μετρική , για να εξετάσουμε την απόδοση των μοντέλων μας.

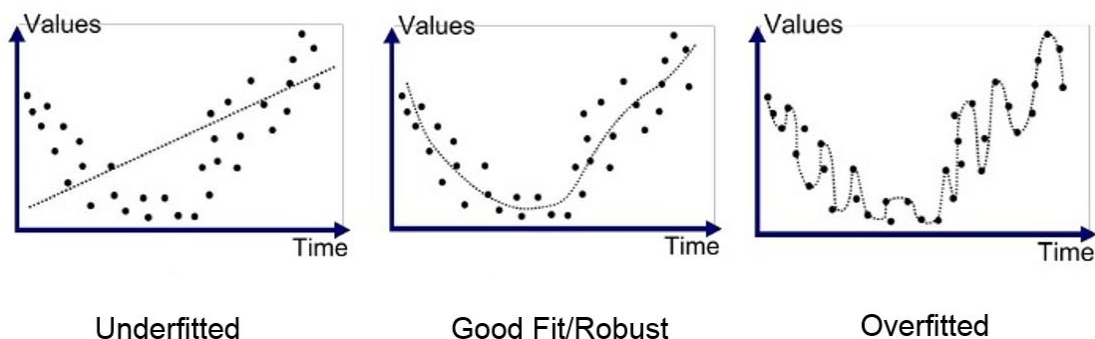
Ένας άλλος διαχωρισμός που αξίζει να σημειώσουμε για τις μετρικές που αναφέρουμε είναι το averaging. Micro και macro averages για οποιεσδήποτε μετρικές υπολογίζουν λίγο διαφορετικά πράγματα και έχουν διαφορετικές ερμηνείες. Μία macro averaged μετρική υπολογίζεται παίρνοντας τον μέσο όρο των μετρικών που αφορούν την κάθε κλάση ξεχωριστά, συνεπώς αντιμετωπίζει την κάθε κλάση ισότιμα. Μία micro averaged μετρική προσθέτει τις συνεισφορές όλων των κλάσεων και υπολογίζει την μετρική στο συνολικό αποτέλεσμα.

Στην περίπτωση μας, θα ασχοληθούμε με macro averaged μετρικές καθώς όπως αναφέραμε και πριν στα σύνολα δεδομένων έχουμε αρκετά μεγάλο imbalance και θέλουμε να δούμε την απόδοση του μοντέλου αντιμετωπίζοντας όλες τις κλάσεις ισότιμα.

Επανερχόμενοι στο στάδιο εκπαίδευσης του νευρωνικού, αξίζει να αναφερθεί ότι αποτελεί καλή πρακτική, και την εφαρμόζουμε και εδώ, όπου μας το επιτρέπει το

πλήθος των δειγμάτων να χρησιμοποιείται πέρα από το training και test set, και ένα validation set. Το validation set είναι ένα σύνολο δεδομένων πάνω στο οποίο εξετάζουμε την απόδοση του νευρωνικού κατά τη διαδικασία της εκπαίδευσης και βοηθάει στον εντοπισμό των κατάλληλων υπερπαραμέτρων του νευρωνικού, όπως είναι παραδείγματος χάρη ο ρυθμός μάθησης (learning rate) αλλά και στην αποφυγή του overfitting.

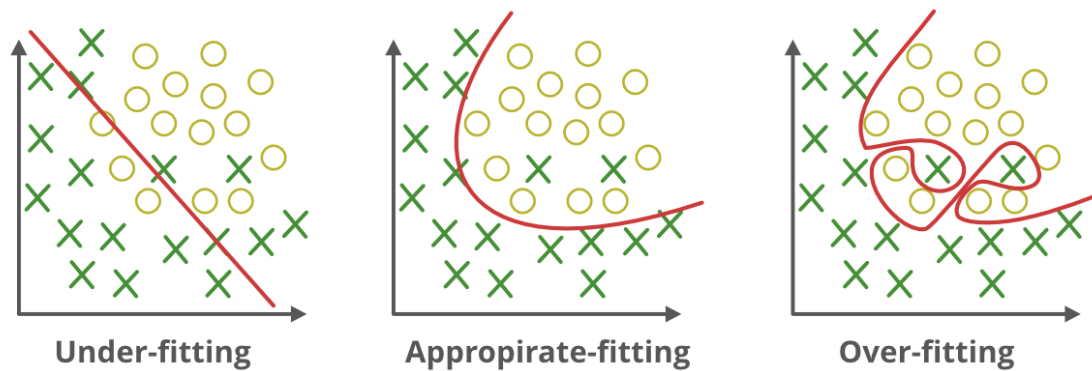
Το overfitting αποτελεί ένα μείζον πρόβλημα στο πεδίο του machine learning και αναφέρεται στην τάση των μοντέλων να μαθαίνουν πολύ καλά τα δεδομένα εκπαίδευσης αλλά να μην γενικεύουν καλά σε δεδομένα που δεν έχουν ξαναδεί. Στην παρακάτω εικόνα, παρουσιάζεται παραστατικά το πρόβλημα του overfitting αλλά και το αντίστροφό του, το underfitting σε ένα πρόβλημα παλινδρόμησης.



Εικόνα 2.5: Overfitting και Underfitting σε πρόβλημα παλινδρόμησης (Πηγή: [17])

Ένα μοντέλο που θα προσπαθούσε να ταιριάζει τα δεδομένα με την πρώτη γραφική παράσταση δεν θα έκανε καλή δουλειά καθώς βλέπουμε ότι πολλά από τα δεδομένα έχουν τιμές αρκετά μακριά από τις προβλεπόμενες, παρατηρούμε δηλαδή underfitting. Αντίθετα, στο overfitting ενώ όλα τα δεδομένα ταιριάζουν απόλυτα με τις προβλεπόμενες τιμές είναι εύκολο να καταλάβουμε ότι το μοντέλο έχει προσπαθήσει να ταιριάζει τις τιμές των δεδομένων σε τέτοιο βαθμό, που δεν έχει μάθει την πραγματική κατανομή τους και δε θα γενικεύσει σωστά. Η χρυσή τομή βρίσκεται στη δεύτερη περίπτωση, όπου παρόλο που τα δεδομένα δεν ταιριάζουν απόλυτα, το μοντέλο έχει μάθει την κατανομή τους αρκετά καλά και είναι σε θέση να γενικεύσει για καινούρια δεδομένα.

Ένα αντίστοιχο παράδειγμα, που αυτή τη φορά αφορά πρόβλημα classification παρουσιάζεται παρακάτω.



Εικόνα 2.6: Overfitting και Underfitting σε πρόβλημα ταξινόμησης (Πηγή: [18])

Εύκολα βλέπουμε ότι και σε αυτή την περίπτωση το κατάλληλο μοντέλο είναι αυτό που κάνει τον μεσαίο διαχωρισμό.

Στην περίπτωση μας, για να αποφύγουμε το overfitting, κατά τη διάρκεια της εκπαίδευσης κρατάμε το μοντέλο που πετυχαίνει το καλύτερο f1 score στο validation set.

## 2.2 Συνελκτικά Νευρωνικά Δίκτυα

Καθώς τα μοντέλα που χρησιμοποιούμε ανήκουν στην κατηγορία των συνελκτικών νευρωνικών δικτύων (CNNs), αξίζει να κάνουμε μία σύντομη αναφορά σε αυτά.

Ένα βασικό πεδίο με το οποίο ασχολείται ο τομέας της μηχανικής μάθησης και στο οποίο έχουν γίνει τεράστια βήματα τα τελευταία χρόνια, είναι το πεδίο της Όρασης Υπολογιστών. Σε αυτό ανήκουν προβλήματα όπως είναι η κατηγοριοποίηση εικόνων ή βίντεο, η εξαγωγή σημασιολογικών χαρακτηριστικών από βίντεο, η αναγνώριση αντικειμένων, αλλά και πολλά άλλα. Ένα μεγάλο μέρος της προόδου που έχει σημειωθεί στο πεδίο αυτό από την μηχανική μάθηση είναι τα συνελκτικά δίκτυα, συχνά αναφερόμενα ως CNNs (Convolutional Neural Networks).

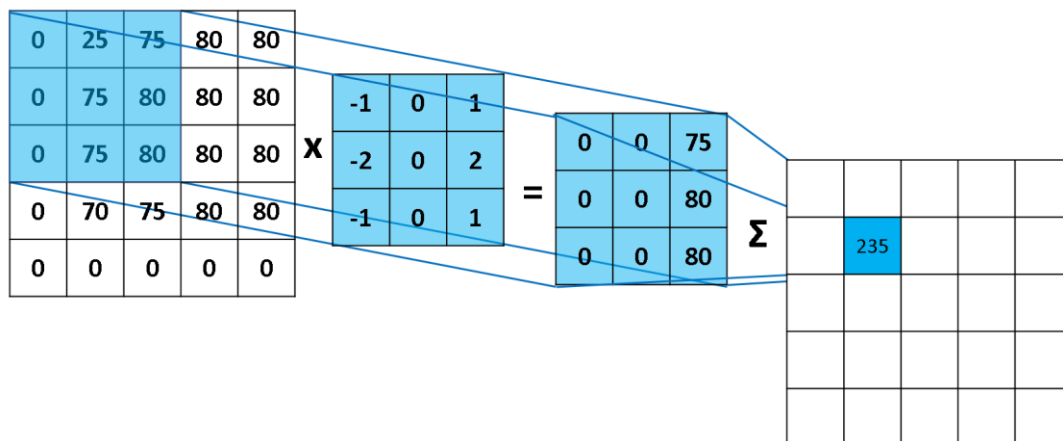
Η αρχιτεκτονική των συνελκτικών δικτύων είναι ανάλογη με το μοτίβο που ακολουθούν οι συνδέσεις νευρώνων στον ανθρώπινο εγκέφαλο και είναι εμπνευσμένη από την οργάνωση του οπτικού φλοιού. Ο κάθε νευρώνας αντιδρά σε εναύσματα που δίνονται μόνο σε μία συγκεκριμένη περιοχή. Μία συλλογή των περιοχών αυτών σχηματίζει αυτό που αναφέρουμε ως οπτικό πεδίο.

Ένα κλασσικό νευρωνικό δίκτυο μπορεί να είχε καλή απόδοση σε προβλήματα που αφορούν απλές εικόνες χαμηλής πολυπλοκότητας. Όταν όμως οι εικόνες γίνονται περισσότερο πολύπλοκες η απόδοση ενός τέτοιου δικτύου πέφτει δραματικά. Τα συνελκτικά νευρωνικά δίκτυα έχουν την ικανότητα να καταλαβαίνουν τις χρονικές και

χωρικές συσχετίσεις σε μία εικόνα εφαρμόζοντας κατάλληλα φίλτρα. Λόγω της μείωσης των παραμέτρων και της επαναχρησιμοποίησης των βαρών η αρχιτεκτονική είναι σε θέση να καταλάβει καλύτερα τις πολυπλοκότητες που μπορεί να κρύβει μία εικόνα.

Αν πάρουμε για παράδειγμα μία εικόνα με τρία κανάλια κόκκινο, πράσινο και μπλε (RGB) με ανάλυση 8K (7680×4320) καταλαβαίνουμε ότι η εκπαίδευση ενός δικτύου με ένα σύνολο δεδομένων που περιέχει ίσως και δεκάδες χιλιάδες από τέτοιες εικόνες μπορεί γρήγορα να γίνει υπολογιστικά ακριβή. Ο ρόλος των συνελκτικών δικτύων είναι να συμπιέζει την εικόνα σε μία μορφή που είναι πιο εύκολο να επεξεργαστεί διατηρώντας ταυτόχρονα τα σημαντικά χαρακτηριστικά της. Έτσι, έχουμε ένα σύστημα ικανό όχι μόνο να εξάγει πολύπλοκα χαρακτηριστικά, αλλά και να κλιμακώνεται σε μεγάλα σύνολα δεδομένων.

Βασικό συστατικό των συνελκτικών δικτύων αποτελεί ο συνελκτικός πυρήνας, η λειτουργία του οποίου αποτυπώνεται στο παρακάτω παράδειγμα.

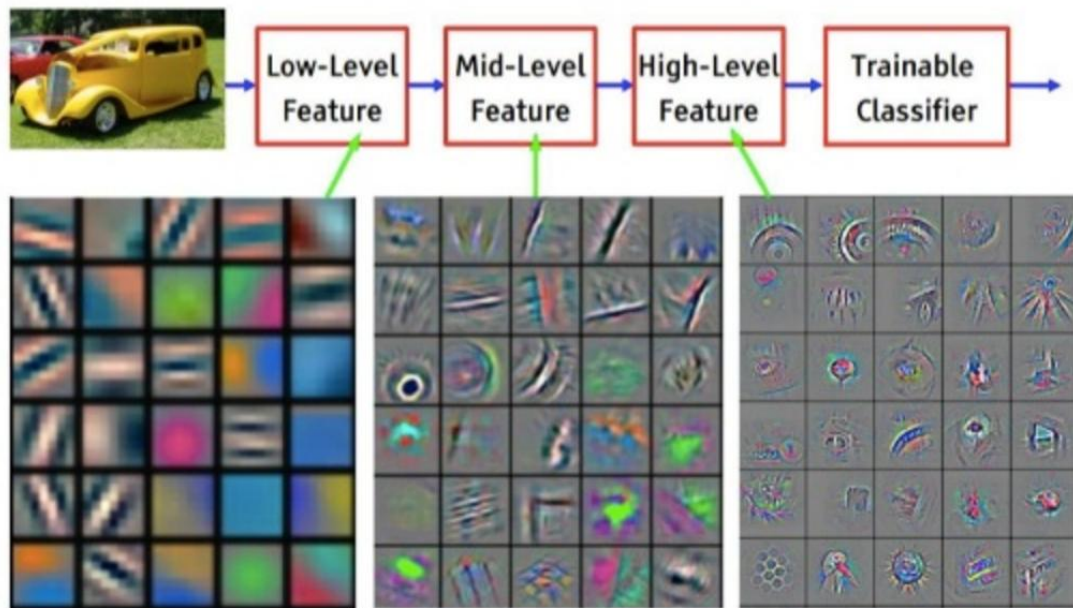


Εικόνα 2.7: Λειτουργία Συνελκτικού Πυρήνα (Πηγή: [19])

Ο πυρήνας μετακινείται στην εικόνα με ένα βήμα που ορίζει ο προγραμματιστής και εφαρμόζει κάθε φορά συνέλιξη με τα pixel της περιοχής που του αντιστοιχούν. Η παραπάνω διαδικασία εφαρμόζεται επαναληπτικά μέχρι να καλυφθεί ολόκληρη η εξεταζόμενη εικόνα.

Όταν έχουμε να κάνουμε με εικόνες με πολλαπλά κανάλια, όπως είναι μία εικόνα RGB, ο πυρήνας έχει το ίδιο βάθος με την εικόνα ( 3 στην περίπτωση RGB ) και όλες οι έξοδοι προστίθενται με τη σταθερά bias για να δώσουν ένα ενιαίο αποτέλεσμα. Ένα συνελκτικό επίπεδο ενός τέτοιου δικτύου δεν περιορίζεται σε έναν μόνο πυρήνα. Αντίθετα, χρησιμοποιούνται αρκετοί πυρήνες ο κάθε ένας υπεύθυνος για την εξαγωγή χαρακτηριστικών διαφορετικού επιπέδου. Για παράδειγμα, τα πρώτα επίπεδα ενός τέτοιου δικτύου μπορεί να είναι υπεύθυνα για την εξαγωγή χαρακτηριστικών όπως είναι οι ακμές ή οι γωνίες ενώ τα τελευταία για την εξαγωγή χαρακτηριστικών υψηλότερου επιπέδου, όπως είναι τα αντικείμενα, παρέχοντας έτσι μία πλήρη κατανόηση του σημασιολογικού

περιεχομένου μίας εικόνας. Ένα παράδειγμα του είδους των χαρακτηριστικών που εξάγει το κάθε επίπεδο δίνεται παρακάτω.



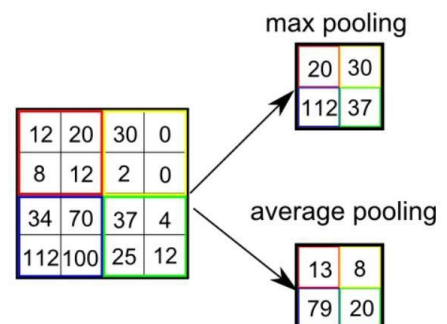
Εικόνα 2.8: Συνελκτικό δίκτυο εκπαιδευμένο στο Imagenet - Κλιμακωτή Εξαγωγή Χαρακτηριστικών (Πηγή: [20])

Ενώ τα πρώτα επίπεδα εξάγουν χαρακτηριστικά όπως οι ακμές τα υψηλότερα επίπεδα εξάγουν χαρακτηριστικά πιο κοντά σε αυτά που παρατηρεί και ένας άνθρωπος όπως οι ρόδες.

Το επόμενο σημαντικό επίπεδο των συνελκτικών δικτύων είναι το επίπεδο pooling. Όπως και το συνελκτικό επίπεδο έτσι και αυτό έχει ως στόχο την μείωση της διάστασης της εισόδου για ταχύτερη επεξεργασία. Επίσης, είναι υπεύθυνο για την διατήρηση των κυρίαρχων χαρακτηριστικών σε μία περιοχή. Το παραπάνω στόχο τον πετυχαίνει με δύο τεχνικές :

- Max pooling και
- Average pooling

Το max pooling επιστρέφει την μέγιστη τιμή από τα pixels της περιοχής που καλύπτει ο πυρήνας ενώ το average pooling την μέση τιμή.

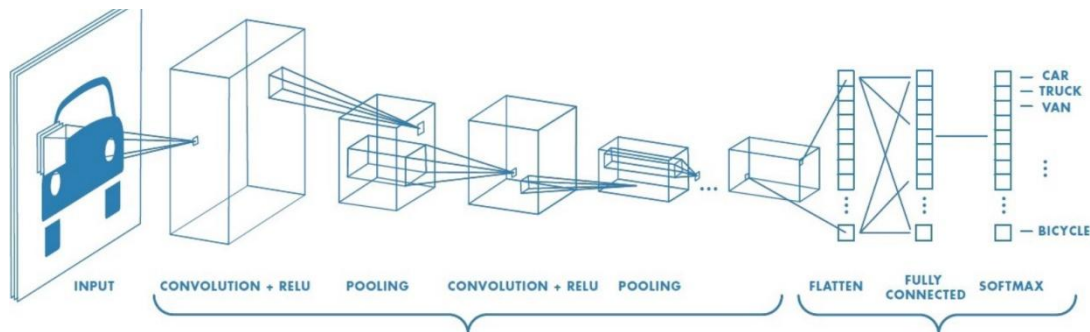


Εικόνα 2.9: Max vs Average Pooling (Πηγή: [21])

Έπειτα από την εξαγωγή των χαρακτηριστικών με την παραπάνω διαδικασία χρησιμοποιούνται fully connected επίπεδα με σκοπό την μετατροπή των εξόδων των προηγούμενων επιπέδων σε μονοδιάστατες αναπαραστάσεις και τελικώς την πρόβλεψη μίας ετικέτας. Το επίπεδο

αυτός είναι ένας εύκολος τρόπος να μάθουμε μη-γραμμικές συναρτήσεις στον χώρο των χαρακτηριστικών.

Η αρχιτεκτονική ενός συνελκτικού δικτύου συνδυάζει τα παραπάνω επίπεδα με έναν τρόπο που φαίνεται στην παρακάτω εικόνα.



Εικόνα 2.10 : Αρχιτεκτονική Συνελκτικού Δικτύου (Πηγή : [21])

Υπάρχουν πολλές γνωστές διαθέσιμες αρχιτεκτονικές συνελκτικών δικτύων. Ένα βασικό δίκτυο το οποίο χρησιμοποιούμε και εμείς στην παρούσα διπλωματική και θα αναλύσουμε στη συνέχεια είναι το ResNet-50.

## 2.3 ResNet-50

Τα residual networks μας επιτρέπουν να εκπαιδεύουμε πολύ βαθιά νευρωνικά δίκτυα ( με 150+ επίπεδα) με μεγάλη επιτυχία. Στα κλασσικά βαθιά νευρωνικά δίκτυα όταν ο αριθμός των επιπέδων αυξάνεται πολύ, παρατηρείται το πρόβλημα των vanishing ή exploding gradients. Κατά την διαδικασία του backpropagation όταν οι παράγωγοι είναι πολύ μικρές ή πολύ μεγάλες, και πολλαπλασιάζονται τότε οι αριθμοί που προκύπτουν είναι είτε σχεδόν μηδενικοί (vanishing) είτε υπερβολικά μεγάλοι (exploding). Έτσι, ενώ θα περιμέναμε ένα δίκτυο με περισσότερα επίπεδα να πετυχαίνει καλύτερη ακρίβεια, κάτι τέτοιο δεν παρατηρείται λόγω των παραπάνω προβλημάτων.

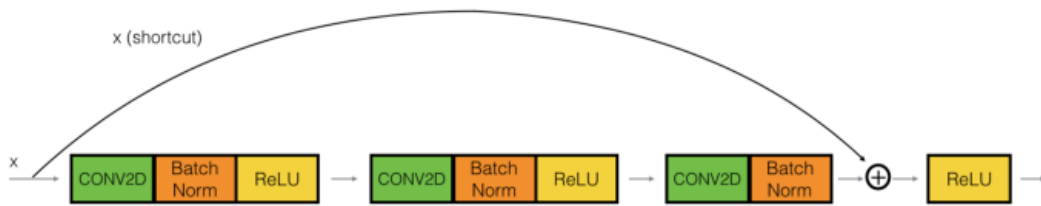
Καταλαβαίνουμε λοιπόν ότι για να δημιουργήσουμε πιο βαθιά νευρωνικά δίκτυα δεν προσθέτουμε απλά επίπεδα το ένα μετά το άλλο. Το ResNet πρώτο εισήγαγε την έννοια του skip connection ( [22] ).



Εικόνα 2.11 : Skip Connection (Πηγή : [23])

Στην αριστερή εικόνα απλά τοποθετούμε επίπεδα το ένα μετά το άλλο. Στη δεξιά εικόνα τοποθετούμε τα επίπεδα το

ένα μετά το άλλο, αλλά προσθέτουμε και την αρχική είσοδο στην έξοδο του επόμενου επιπέδου. Παρακάτω φαίνεται ένα παράδειγμα εφαρμογής της παραπάνω τεχνικής σε ένα τμήμα ενός συνελκτικού δικτύου.

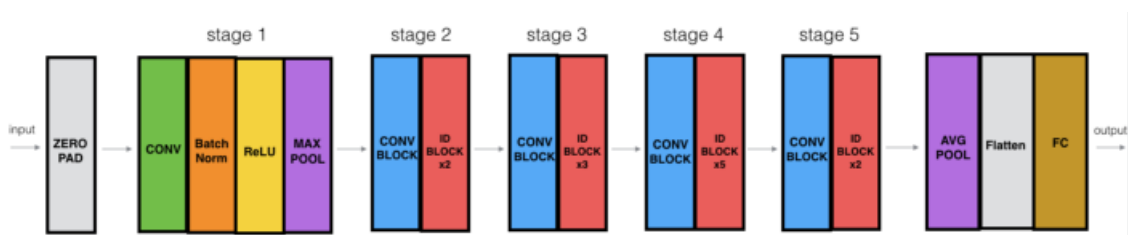


Εικόνα 2.12 : Skip Connection σε CNN (Πηγή : [23])

Το skip connection αξίζει να σημειωθεί ότι εφαρμόζεται πριν την συνάρτηση ενεργοποίησης, καθώς έχει βρεθεί ότι με τον τρόπο αυτό παίρνουμε καλύτερα αποτελέσματα στην πράξη.

Το ResNet-50 αποτελείται από 5 στάδια το κάθε ένα με ένα συνελκτικό και ένα ταυτοτικό block. Κάθε συνελκτικό block έχει 3 συνελκτικά επίπεδα και κάθε ταυτοτικό block έχει επίσης 3 συνελκτικά επίπεδα. Συνολικά, το ResNet-50 διαθέτει πάνω από 23 εκατομμύρια παραμέτρους.

Η αρχιτεκτονική του φαίνεται παρακάτω.



Εικόνα 2.13 : ResNet - 50 (Πηγή : [23])

Το μοντέλο αυτό το χρησιμοποιούμε ως βάση για τα μοντέλα που χρησιμοποιούμε στα πειράματα που θα ακολουθήσουν. Παρόλα αυτά δεν χρησιμοποιούμε το μοντέλο με τυχαία αρχικοποιημένες παραμέτρους. Αντίθετα, εφαρμόζουμε μία ευρέως χρησιμοποιούμενη τεχνική στον τομέα της βαθιάς μάθησης, το transfer learning. Την τεχνική αυτή αναλύουμε συνοπτικά στη συνέχεια.

## 2.4 Μεταφορά Μάθησης (Transfer Learning)

Σαν άνθρωποι, όταν προσπαθούμε να μάθουμε μία καινούρια δεξιότητα δεν ξεκινάμε πάντα από το μηδέν. Αντίθετα, προσπαθούμε να αντλήσουμε γνώσεις από την προηγούμενη εμπειρία μας και να τις προσαρμόσουμε ώστε να είναι χρήσιμες για τη λύση ενός καινούριου προβλήματος. Για παράδειγμα, ένας άνθρωπος που γνωρίζει ιταλικά και θέλει να μάθει γαλλικά είναι προφανώς σε πιο ευνοϊκή θέση από κάποιον που δεν γνωρίζει

καμία από τις δύο γλώσσες, καθώς αυτές, όντας λατινογενείς, έχουν πολλά κοινά μεταξύ τους, τα οποία μπορεί να αξιοποιήσει κανείς για να μάθει πιο εύκολα την καινούρια γλώσσα.

Όπως οι άνθρωποι αξιοποιούμε ήδη αποκτημένες γνώσεις, όταν προσπαθούμε να μάθουμε κάτι καινούριο, έτσι και στον τομέα της βαθιάς μάθησης, όταν εκπαιδεύουμε ένα μοντέλο για να λύσει ένα συγκεκριμένο πρόβλημα, δεν είναι ανάγκη να ξεκινάμε πάντα από το μηδέν. Μπορούμε να αξιοποιήσουμε γνώσεις που έχουμε πάρει από την επίλυση άλλων προβλημάτων. Η τεχνική αυτή ονομάζεται transfer learning ([11], [24], [25]).

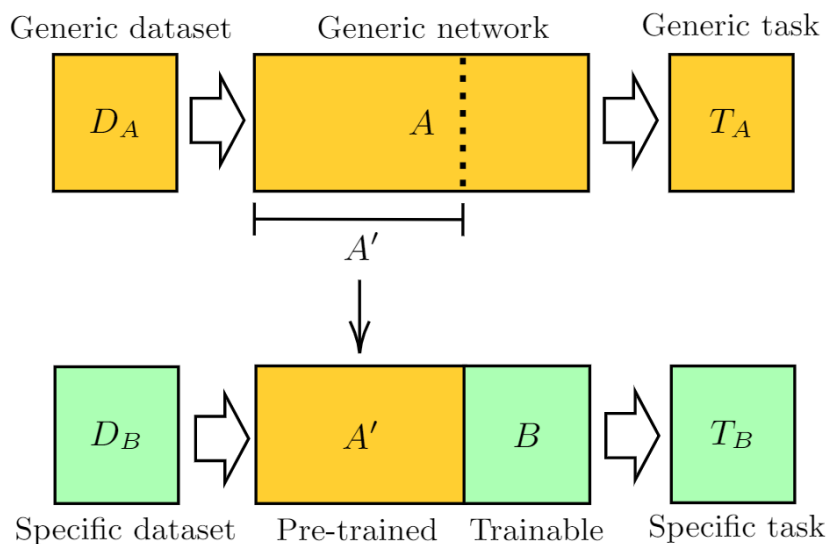
Το transfer learning έχει πολλά πλεονεκτήματα. Όπως έχουμε πει για να εκπαιδεύσουμε ένα μοντέλο χρειαζόμαστε τεράστιο όγκο δεδομένων. Τέτοια σύνολα δεδομένων όμως, είναι αρκετά δύσκολο να βρεθούν για όλα τα προβλήματα. Στην περίπτωσή μας οι ακτινογραφίες και οι αξονικές τομογραφίες ασθενών ήταν αρκετά περιορισμένες, κάτι το οποίο καθιστά δύσκολη την επιτυχή εκπαίδευση ενός δικτύου. Το transfer learning μπορεί να μας βοηθήσει να επιλύσουμε αυτό το πρόβλημα.

Επίσης, ακόμα και αν είχαμε μεγάλο όγκο δεδομένων και μπορούσαμε να εκπαιδεύσουμε ένα μοντέλο επιτυχώς πάνω στα δεδομένα αυτά, ελλοχεύει πάντα ο κίνδυνος του overfitting. Δηλαδή, το μοντέλο να πετυχαίνει καλή απόδοση στα δεδομένα αυτά, αλλά να μην γενικεύει καλά σε νέα δεδομένα. Το transfer learning μας βοηθάει και σε αυτή την περίπτωση.

Στην πραγματικότητα, υπάρχουν αρκετές διαφορετικές εκδοχές του transfer learning. Εμείς θα ασχοληθούμε με την περίπτωση όπου έχουμε επισημειώσεις τόσο για τις εισόδους του αρχικού προβλήματος, όσο και για τις εισόδους του προβλήματος που θέλουμε να αντιμετωπίσουμε. Συγκεκριμένα, ως βάση για τα μοντέλα μας χρησιμοποιούμε επίπεδα από το ResNet-50 όπως αναφέραμε και προηγουμένως. Τα βάρη των επιπέδων αυτών δεν αρχικοποιούνται τυχαία. Χρησιμοποιούμε βάρη τα οποία έχει αποκτήσει το δίκτυο κατά την εκπαίδευσή του πάνω στο Imagenet.

Το Imagenet είναι μία τεράστια βάση δεδομένων με πάνω από 14 εκατομμύρια επισημειωμένες εικόνες, που ανήκουν σε πάνω από 20 χιλιάδες κατηγορίες. Εκπαιδεύοντας το δίκτυο αρχικά σε αυτό το σύνολο δεδομένων, τα πρώτα επίπεδα, που όπως αναφέραμε είναι υπεύθυνα για την εξαγωγή χαμηλού επιπέδου χαρακτηριστικών μαθαίνουν πολύ καλά να αναγνωρίζουν χαρακτηριστικά όπως είναι οι ακμές και οι γωνίες. Στη συνέχεια, συμπληρώνουμε το δίκτυο με έναν αριθμό από fully connected επίπεδα και επανεκπαιδεύουμε χρησιμοποιώντας αυτή τη φορά δεδομένα από το δικό μας πρόβλημα, δηλαδή ακτινογραφίες ή αξονικές τομογραφίες αντίστοιχα. Χονδρικά, η διαδικασία που ακολουθούμε φαίνεται στο παρακάτω σχήμα.





Εικόνα 2.14 : Transfer Learning (Πηγή : [26])

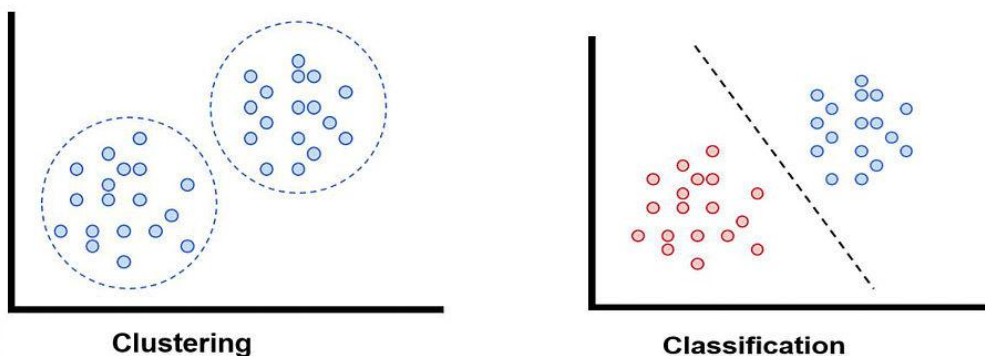
Εφαρμόζοντας την παραπάνω τεχνική, όχι μόνο έχουμε καλύτερη απόδοση στο πρόβλημα που αντιμετωπίζουμε, αλλά δημιουργούμε και ένα μοντέλο που είναι σε θέση να γενικεύει καλύτερα, αφού συνδυάζει γνώσεις από διαφορετικά προβλήματα. Προστατευόμαστε έτσι με έναν επιπλέον τρόπο από το overfitting.

Πριν προχωρήσουμε στα πειράματα ακολουθεί μία συνοπτική παρουσίαση της ομαδοποίησης (clustering) καθώς είναι μία έννοια που θα μας απασχολήσει στη συνέχεια.

## 2.5 Συσταδοποίηση (Clustering)

Μέχρι τώρα οι αλγόριθμοι που έχουμε εξετάσει αφορούν περιπτώσεις επιβλεπόμενης μάθησης. Στην επιβλεπόμενη μάθηση, η μάθηση επιτυγχάνεται χρησιμοποιώντας τις πραγματικές τιμές των δεδομένων εκπαίδευσης. Ο σκοπός είναι το μοντέλο να μάθει να προβλέπει μία κατηγορία για ένα παράδειγμα, από ένα σύνολο προκαθορισμένων διαθέσιμων κατηγοριών. Στην μη-επιβλεπόμενη μάθηση, το μοντέλο προσπαθεί να μάθει τη φυσική δομή που διέπει ένα σύνολο από μη-επισημασμένα διαθέσιμα δεδομένα.

Η ομαδοποίηση (clustering) είναι ένας τρόπος μη-επιβλεπόμενης μάθησης. Κατά το clustering προσπαθούμε να οργανώσουμε τα δεδομένα μας σε πυρήνες (clusters). Τι είναι όμως ένας πυρήνας; Απλοϊκά μπορούμε να πούμε ότι είναι ένα σύνολο δεδομένων που μοιάζουν μεταξύ τους και διαφέρουν από αντικείμενα που ανήκουν σε διαφορετικούς πυρήνες. Η διαφορά ανάμεσα στην ομαδοποίηση και την κατηγοριοποίηση αποτυπώνεται καλύτερα στην παρακάτω εικόνα.



Εικόνα 2.15 : Transfer Learning (Πηγή : [27])

Υπάρχουν πολλοί διαφορετικοί αλγόριθμοι για να πετύχουμε το clustering, ο κάθε ένας από τους οποίους χρησιμοποιεί διαφορετικές τεχνικές για να σχηματίσει τον κάθε cluster. Εμείς θα χρησιμοποιήσουμε τον αλγόριθμο K-Means ([28]).

Ο αλγόριθμος K-Means χωρίζει τα δεδομένα σε K ομάδες. Επιστρέφει τα κέντρα των ομάδων αυτών και μία ετικέτα για το κάθε δεδομένο εκπαίδευσης (κάθε δεδομένο αντιστοιχίζεται σε έναν μοναδικό cluster). Ο αλγόριθμος αποτελείται από δύο βήματα που φαίνονται στην παρακάτω εικόνα.

Για

- ένα σύνολο στοιχείων  $D = \{x_1, x_2, x_3, \dots, x_n\}$  και
- έναν επιθυμητό αριθμό clusters  $k$

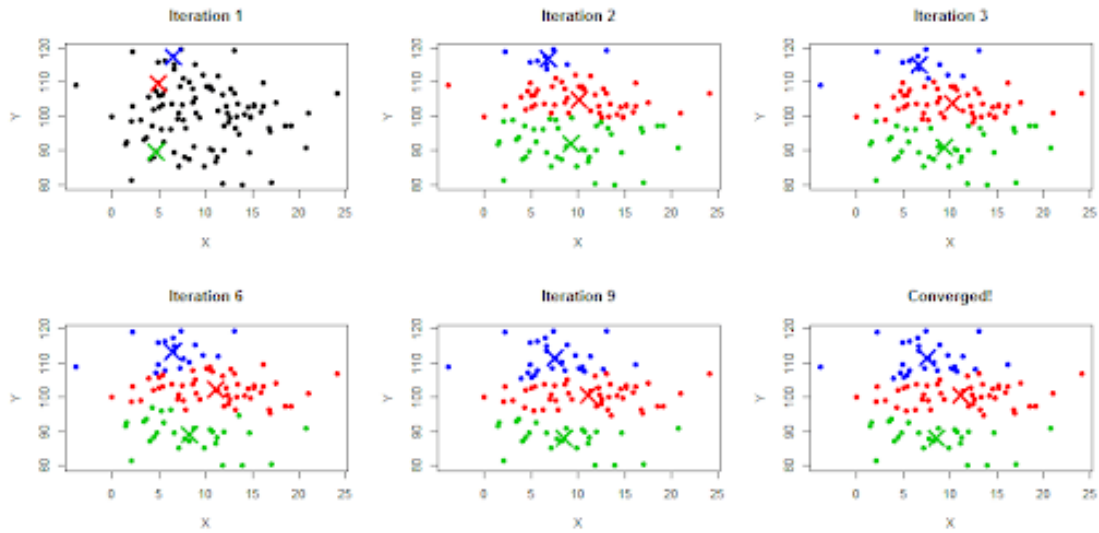
Αναθέτουμε αρχικά τυχαίες τιμές στα κέντρα  $c_1, c_2, \dots, c_k$ .

Έπειτα εφαρμόζουμε επαναληπτικά:

- Αναθέτουμε κάθε  $x_i$  στο cluster με του οποίου το κέντρο η απόσταση είναι η μικρότερη.
- Υπολογίζουμε νέα κέντρα για κάθε cluster. Συγκεκριμένα, θέτουμε κάθε κέντρο ίσο με τη μέση τιμή των στοιχείων που ανήκουν στο cluster.

Η παραπάνω διαδικασία επαναλαμβάνεται μέχρι να ικανοποιηθεί ένα κριτήριο σύγκλισης που θέτουμε.

Στην παρακάτω εικόνα ακολουθεί ένα παράδειγμα εφαρμογής του παραπάνω αλγορίθμου.



Εικόνα 2.16 : Παράδειγμα K-Means Clustering (Πηγή : [29])

Έχοντας καλύψει πλήρως τις θεωρητικές έννοιες που θα χρειαστούμε προχωράμε στα αποτελέσματα της εκτέλεσης των πειραμάτων, ξεκινώντας από αυτά που αφορούν σύνολα δεδομένων με ακτινογραφίες.

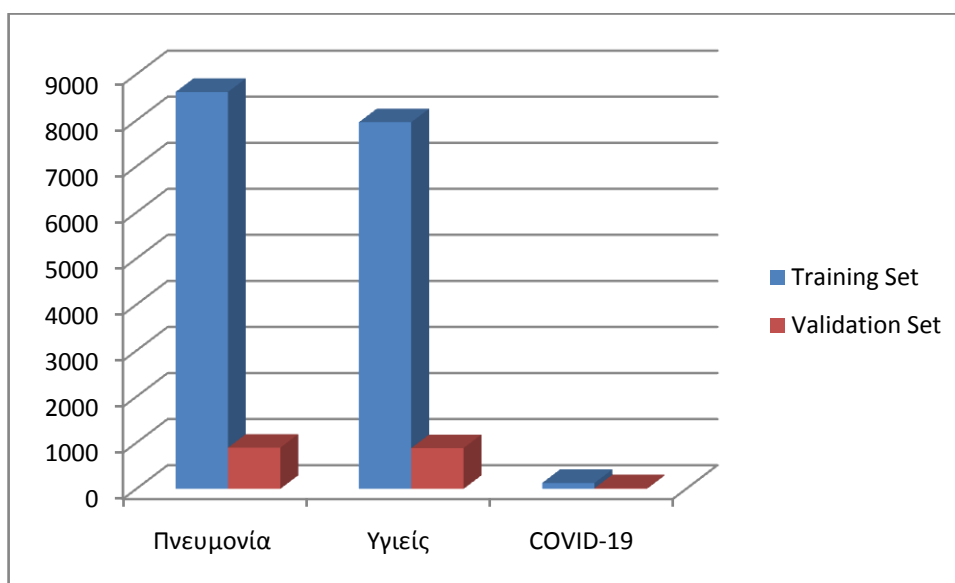
### 3. Σύνολο Δεδομένων με Ακτινογραφίες

#### 3.1 Covid-Next

Το δίκτυο που χρησιμοποιήσαμε έχει αρχικά 5 blocks που περιέχουν layers του ResNet-50. Ακολουθούνται από 3 συνελκτικά και batch normalization επίπεδα και τέλος ένα fully connected επίπεδο με αριθμό εξόδων που προσδιορίζουμε εμείς ανάλογα με το πρόβλημα που καλούμαστε να αντιμετωπίσουμε κάθε φορά.

#### 3.2 Chest X-Ray Dataset

Το πρώτο σύνολο δεδομένων στο οποίο θα δοκιμάσουμε το δίκτυο μας αποτελεί μία σύμπτυξη των συνόλων δεδομένων COVID Chest Xray dataset που βρίσκεται στη διεύθυνση [30] και τα σύνολο δεδομένων Kaggle RSNA Pneumonia dataset που βρίσκεται στη διεύθυνση [31]. Χωρίζουμε τα δεδομένα σε training και validation sets, καθώς λόγω έλλειψης δειγμάτων με κορονοϊό τα δεδομένα είναι πολύ λίγα για να χρησιμοποιήσουμε και test set. Έτσι, έχουμε training set και validation set με τις εξής κατανομές.



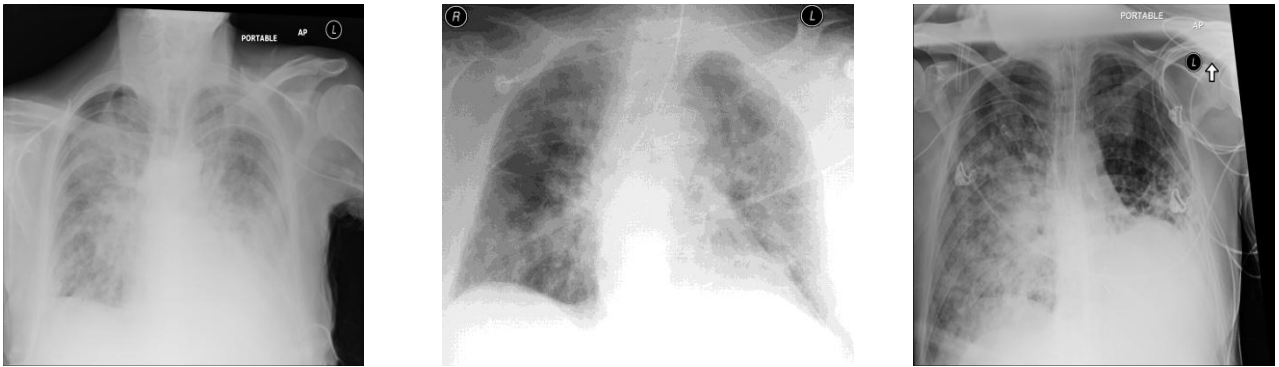
Εικόνα 3.1 : Κατανομές training και validation sets

Συγκριμένα έχουμε

- στο training set 7960 ακτινογραφίες υγιών ανθρώπων, 8617 ακτινογραφίες ασθενών με πνευμονία και 128 ακτινογραφίες ασθενών με COVID-19
- στο validation set 885 ακτινογραφίες υγιών ανθρώπων, 893 ακτινογραφίες ασθενών με πνευμονία και 14 ακτινογραφίες ασθενών με COVID-19

Παρακάτω παρουσιάζουμε ενδεικτικά κάποιες ακτινογραφίες από την κάθε κλάση.

- Πνευμονία



Εικόνα 3.2 : Ακτινογραφίες Στήθους Ασθενών με Πνευμονία

- Υγιείς



Εικόνα 3.3 : Ακτινογραφίες Στήθους Υγιών Ανθρώπων

- COVID-19



Εικόνα 3.4 : Ακτινογραφίες Στήθους Ασθενών με COVID-19

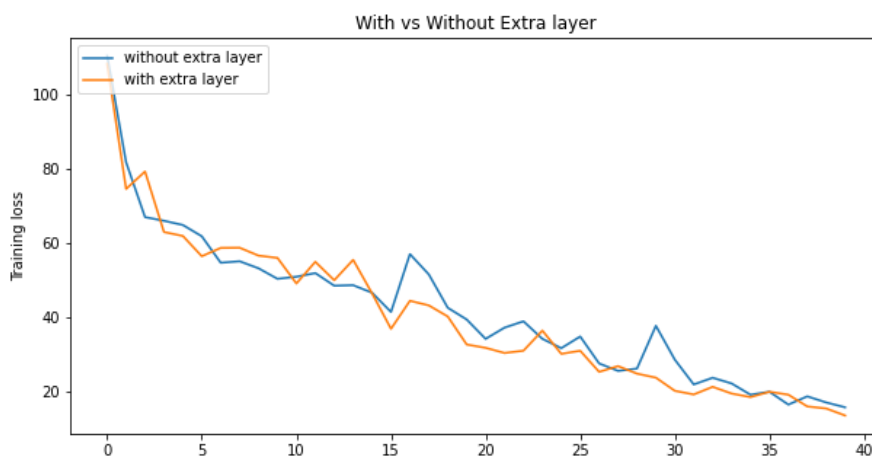
Επειδή οι ακτινογραφίες είχαν αρκετά μεγάλο μέγεθος για να τις δώσουμε ως είσοδο στο νευρωνικό δίκτυο τις μετασηματίσαμε σε διαστάσεις 256x256.

Κατά τη διάρκεια της εκπαίδευσης για να αποφύγουμε το νευρωνικό δίκτυο να πάθει overtraining εφαρμόσαμε και άλλους μετασχηματισμούς στις εικόνες με τυχαίο τρόπο. Συγκεκριμένα οι μετασχηματισμοί που εφαρμόσαμε:

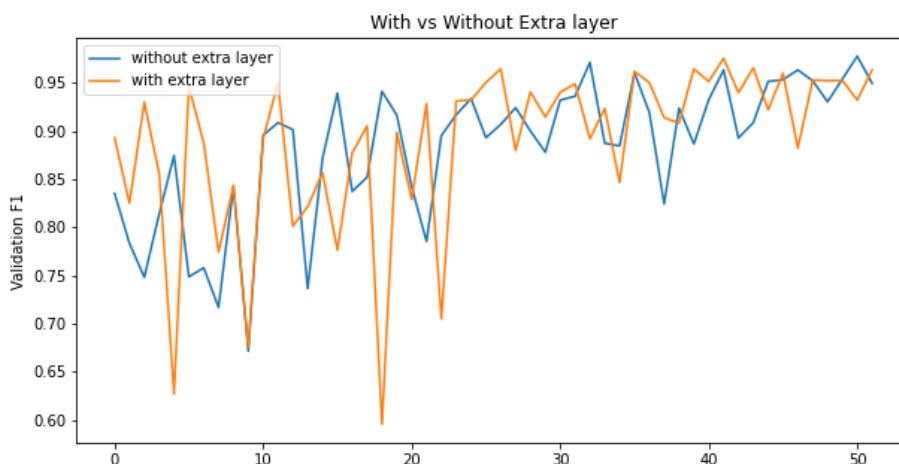
- Για το training : Resize σε 256x256 , τυχαίο κάθετο και οριζόντιο flip με πιθανότητα 0.5, random Affine με πιθανότητα 0.5, color Jitter με πιθανότητα 0.5.
- Για το testing : Resize σε 256x256.

### 3.3 Χρήση Επιπλέον Επιπέδου (With vs Without Extra layer)

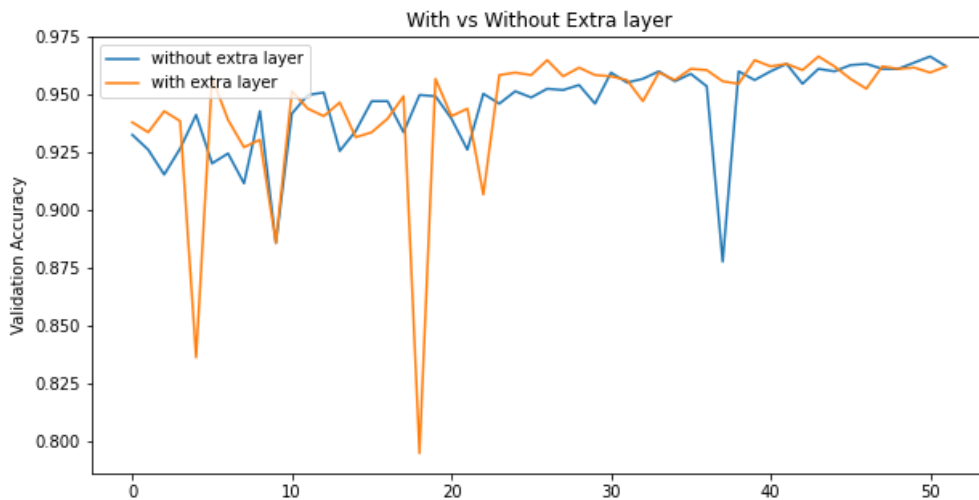
Καθώς ένα μεγάλο κομμάτι της διπλωματικής αφορά το clustering των δεδομένων θα ήταν χρήσιμο να εισάγουμε ένα επιπλέον κρυμμένο επίπεδο στο μοντέλο, το οποίο θα είχε έξοδο διάστασης κατάλληλης για clustering. Αυτή τη στιγμή το προτελευταίο επίπεδο έχει έξοδο διάστασης 512. Εισάγουμε λοιπόν ένα επίπεδο με διάσταση 32 και συγκρίνουμε την απόδοση των δύο μοντέλων, με και χωρίς το επιπλέον επίπεδο. Τα αποτελέσματα της σύγκρισης φαίνονται στα παρακάτω γραφήματα. Όπως αναφέραμε και σε προηγούμενο κεφάλαιο η μετρική που θα μας απασχολήσει κυρίως είναι το F1 score, καθώς δεν επηρεάζεται ιδιαίτερα από άνισες κατανομές στις κλάσεις του συνόλου δεδομένων όπως παρατηρείται εδώ. Παρουσιάζουμε, γραφήματα και για το training loss και την accuracy στα δεδομένα του validation set όπου στον οριζόντιο άξονα απεικονίζεται η εποχή .



Εικόνα 3.5 : Training Loss with and without extra layer



Εικόνα 3.6 : Validation F1 with and without extra layer

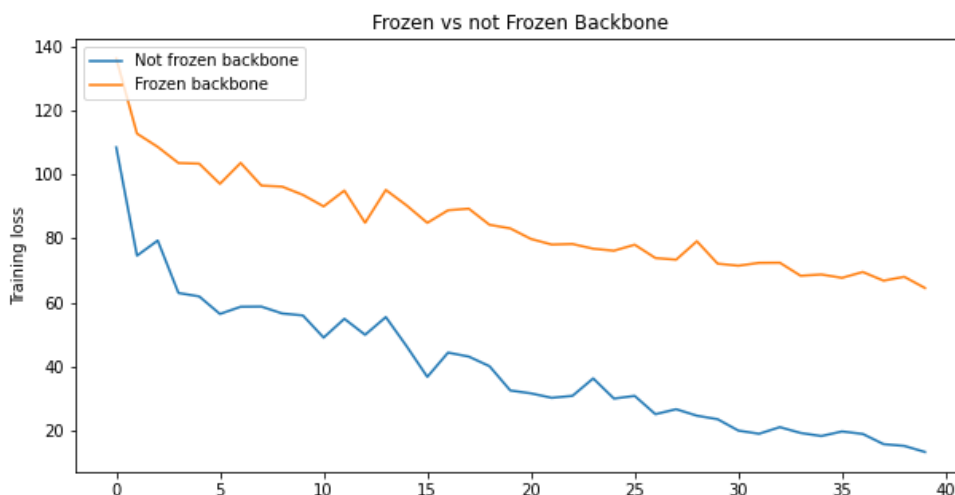


Εικόνα 3.7 : Validation Accuracy with and without extra layer

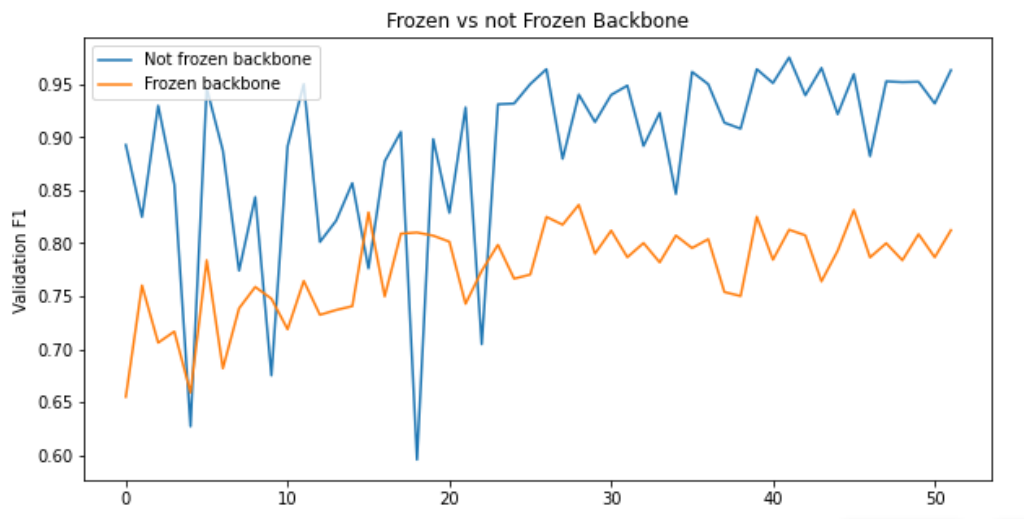
Όπως μπορούμε εύκολα να συμπεράνουμε από τα παραπάνω γραφήματα εισάγοντας ένα επιπλέον επίπεδο δεν έχουμε μεγάλη αλλαγή στην απόδοση. Έχουμε όμως αρκετά επιπρόσθετα οφέλη, όπως θα δούμε στη συνέχεια όταν επιχειρήσουμε να εφαρμόσουμε ομαδοποίηση στα δεδομένα μας. Έτσι, στη συνέχεια τα πειράματα που θα εξετάσουμε θα αφορούν μοντέλα που έχουν το επιπλέον επίπεδο.

### 3.4 Επανεκπαίδευση ή μη της Βασικής Δομής (Frozen vs Not Frozen Backbone)

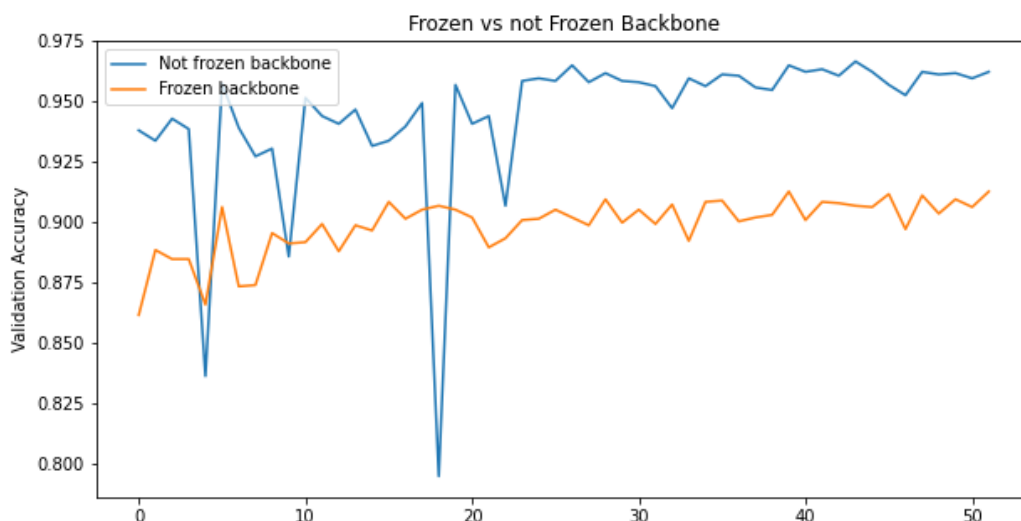
Μία τεχνική που χρησιμοποιείται ευρέως στο transfer learning είναι αντί να εκπαιδεύουμε ολόκληρο το pretrained δίκτυο στο σύνολο δεδομένων του προβλήματός μας, να κρατάμε "παγωμένα" κάποια επίπεδα, χωρίς να αφήνουμε τα βάρη τους να ανανεωθούν κατά την εκπαίδευση και να εκπαιδεύουμε μόνο έναν αριθμό από τα τελευταία επίπεδα. Θα δοκιμάσουμε αυτή την τεχνική και στο δικό μας μοντέλο. Κρατάμε παγωμένα τα πρώτα 5 blocks που περιέχουν επίπεδα του ResNet-50 εκπαιδευμένο στο Imagenet και αφήνουμε τα υπόλοιπα να εκπαιδευτούν ελεύθερα. Παρακάτω παρουσιάζουμε τα αποτελέσματα.



Εικόνα 3.8 : Training loss with frozen and not frozen backbone



Εικόνα 3.8 : Validation F1 with frozen and not frozen backbone



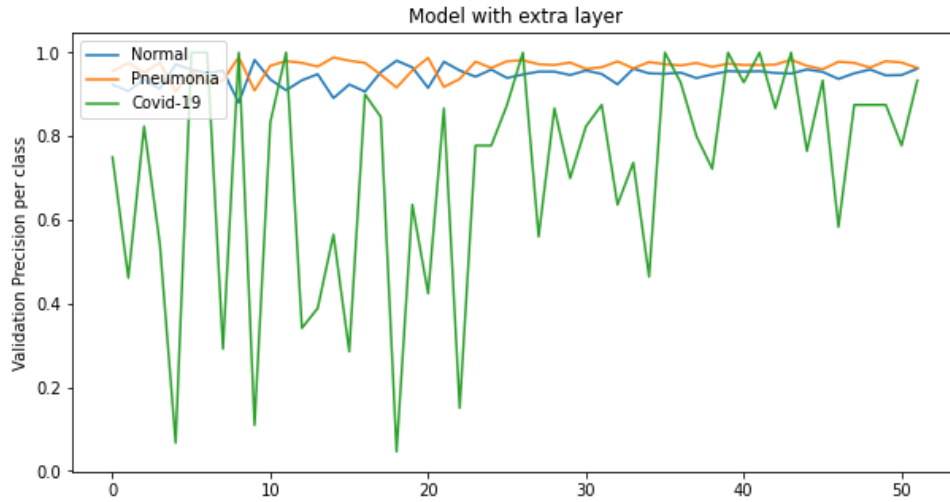
Εικόνα 3.9 : Validation Accuracy with frozen and not frozen backbone

Όπως βλέπουμε εύκολα, εξετάζοντας οποιαδήποτε από τις 3 μετρικές, αυτός ο τρόπος εφαρμογής του transfer learning δεν πετυχαίνει καλά αποτελέσματα στην περίπτωση μας, οπότε δε θα τον εφαρμόσουμε.

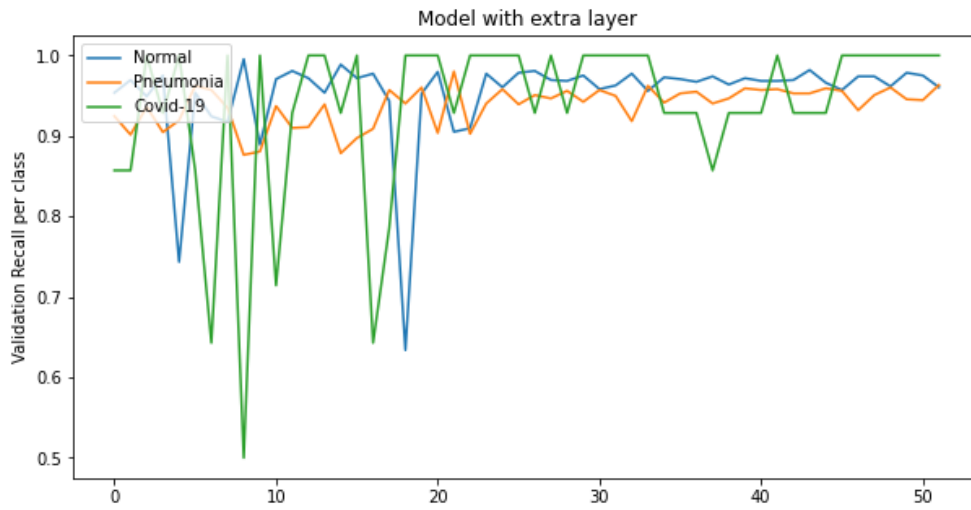
### 3.5 Μετρικές ανά κλάση

Έχοντας καταλήξει στην αρχιτεκτονική του μοντέλου που θα χρησιμοποιήσουμε αλλά και στον τρόπο με τον οποίο θα εφαρμόσουμε το transfer learning έχει ενδιαφέρον να εξετάσουμε την απόδοση του μοντέλου μας σε κάθε κλάση ξεχωριστά. Να σημειώσουμε ότι οι μετρικές που έχουμε αναφέρει μέχρι τώρα είναι macro averaged. Παρακάτω φαίνονται τα γραφήματα 3 μετρικών (precision, recall και f1 score) στο validation set για κάθε μία από τις 3 κλάσεις κατά τη διάρκεια της εκπαίδευσης.

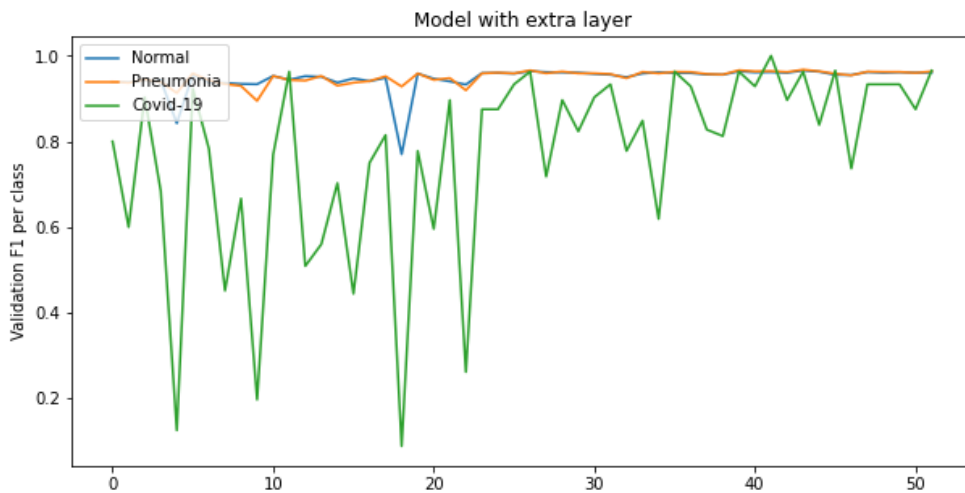




Εικόνα 3.10 : Precision per class on validation set



Εικόνα 3.11 : Recall per class on validation set



Εικόνα 3.12 : F1 score per class on validation set

Όπως παρατηρούμε από τα παραπάνω διαγράμματα έχουμε λίγο χειρότερη απόδοση στις μετρικές που αφορούν την κλάση COVID-19. Κάτι τέτοιο βέβαια είναι αναμενόμενο αφού διαθέτουμε τάξεις μεγέθους (εκατονταπλάσια) περισσότερα δεδομένα για τις άλλες δύο κλάσεις.

### 3.6 Συσταδοποίηση σε 3 Κλάσεις

Το επόμενο πείραμα που θα πραγματοποιήσουμε αποσκοπεί στην εφαρμογή clustering στα δεδομένα μας και στην οπτικοποίηση των αποτελεσμάτων. Για να το πετύχουμε αυτό θα χρησιμοποιήσουμε ως βάση το μοντέλο με το επιπλέον επίπεδο με τους 32 νευρώνες. Η διαδικασία που θα εφαρμόσουμε έχει ως εξής:

- Αρχικά, εκπαιδύουμε ένα τέτοιο δίκτυο στα δεδομένα εκπαίδευσης, χρησιμοποιώντας τους κατάλληλους μετασχηματισμούς που αναφέραμε και προηγουμένως.
- Έπειτα, αφαιρούμε το τελευταίο επίπεδο. Έτσι, η διάσταση της εξόδου τώρα θα είναι 32.
- Περνάμε μία φορά όλα τα δεδομένα εκπαίδευσης από το δίκτυο αποθηκεύοντας τις εξόδους.
- Εφαρμόζουμε k-means clustering στις εξόδους με αριθμό πυρήνων k.
- Για να βρούμε το k δοκιμάζουμε διάφορες τιμές, αναθέτοντας μία ετικέτα (υγιείς, πνευμονία ή COVID-19) στον κάθε cluster και κρατώντας το k για το οποίο παίρνουμε το καλύτερο accuracy και f1 score. Για τα συγκεκριμένα δεδομένα, προκύπτει ότι ο κατάλληλος αριθμός πυρήνων είναι 5.
- Τέλος, οπτικοποιούμε τα κέντρα των πυρήνων. Για να το πετύχουμε αυτό δοκιμάζουμε για κάθε ακτινογραφία των δεδομένων εκπαίδευσης, ποια απέχει την ελάχιστη ευκλείδεια απόσταση από το κάθε κέντρο. Οι εικόνες με την ελάχιστη απόσταση είναι αυτές που απεικονίζουμε ως κέντρα των πυρήνων.

Στη συνέχεια, παρουσιάζουμε τα αποτελέσματα της παραπάνω διαδικασίας συγκριτικά με την κλασική κατηγοριοποίηση, υπό την μορφή classification reports. Τα classification reports είναι μία χρήσιμη συνάρτηση της βιβλιοθήκης scikit-learn που παρουσιάζουν συνοπτικά όλες τις βασικές μετρικές που θέλουμε, τόσο για κάθε κλάση ξεχωριστά όσο και macro-averaged.

Αρχικά, δίνουμε το classification report για τις προβλέψεις που γίνονται χρησιμοποιώντας ομαδοποίηση στα δεδομένα του συνόλου εκπαίδευσης.

	precision	recall	F1-score	support
Υγιείς	0.99	0.99	0.99	7960
Πνευμονία	0.99	0.99	0.99	8617
COVID-19	1.00	1.00	1.00	128
accuracy			0.99	
Macro average	0.99	0.99	0.99	
Weighted average	0.99	0.99	0.99	

Εικόνα 3.13 : Classification Report - Clustering σε δεδομένα εκπαίδευσης ακτινογραφιών

Παρατηρούμε ότι πετυχαίνουμε πολύ καλά αποτελέσματα στα δεδομένα εκπαίδευσης με F1 score 0.99.

Δοκιμάζουμε την τεχνική αυτή και στα δεδομένα που ανήκουν στο validation set.

	precision	recall	F1-score	support
Υγιείς	0.97	0.95	0.96	885
Πνευμονία	0.95	0.97	0.96	893
COVID-19	1.00	0.93	0.96	14
accuracy			0.96	1792
Macro average	0.97	0.95	0.96	1792
Weighted average	0.96	0.96	0.96	1792

Εικόνα 3.14 : Classification Report - Clustering σε δεδομένα επαλήθευσης ακτινογραφιών

Όπως ήταν αναμενόμενο η απόδοση πέφτει όταν δοκιμάζουμε την τεχνική αυτή στα δεδομένα του validation set. Παραμένει όμως πολύ ικανοποιητική.

Για να έχουμε ένα μέτρο σύγκρισης παρακάτω δίνουμε το classification report στο validation set του μοντέλου που δεν χρησιμοποιεί clustering.

	precision	recall	F1-score	support
Υγιείς	0.96	0.97	0.96	885
Πνευμονία	0.97	0.96	0.96	893
COVID-19	1.00	1.00	1.00	14
accuracy			0.96	1792
Macro average	0.98	0.98	0.98	1792
Weighted average	0.96	0.96	0.96	1792

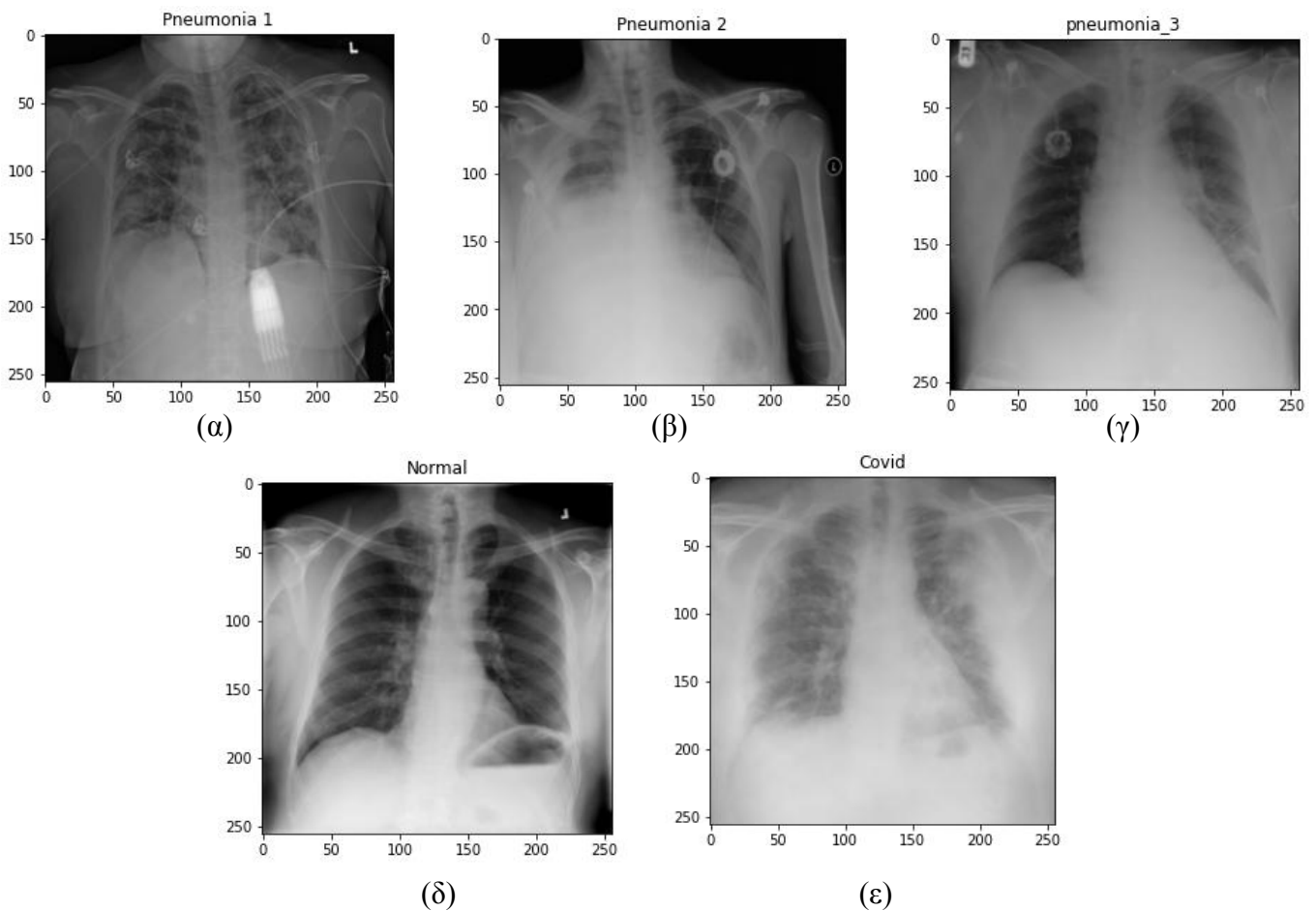
Εικόνα 3.15 : Classification Report - Χωρίς Clustering σε δεδομένα επαλήθευσης ακτινογραφιών

Με τη μέθοδο του clustering παίρνουμε ελαφρώς χειρότερα αποτελέσματα αλλά έχουμε αρκετά επιπρόσθετα οφέλη. Συγκεκριμένα, όπως είπαμε και πριν το clustering μπορεί να μας βοηθήσει να αντιμετωπίσουμε το πρόβλημα του "μαύρου κουτιού" που πολύ συχνά συναντάται σε εφαρμογές μηχανικής μάθησης. Ιδιαίτερα σε συστήματα που έχουν να κάνουν με την παροχή βοήθειας σε ιατρικό προσωπικό κατά τη λήψη αποφάσεων είναι

σημαντικό να υπάρχει διαφάνεια όσον αφορά τις προβλέψεις που κάνει το νευρωνικό δίκτυο.

Χρησιμοποιώντας clustering, για να κάνουμε μία πρόβλεψη για μία εικόνα εξετάζουμε ουσιαστικά την ετικέτα που έχει το κέντρο πυρήνα στο οποίο βρίσκεται πιο κοντά η εικόνα. Οπτικοποιώντας το κέντρο αυτό μπορούμε να βοηθήσουμε το ιατρικό προσωπικό να καταλάβει τι οδήγησε το νευρωνικό να πάρει μία συγκεκριμένη απόφαση, εξετάζοντας τις ομοιότητες ανάμεσα στις δύο εικόνες.

Έτσι, εφαρμόζοντας την τεχνική για την οποία μιλήσαμε στην αρχή του κεφαλαίου παρέχουμε παρακάτω τις ακτινογραφίες οι οποίες αντιστοιχούν στα 5 κέντρα που προέκυψαν.



Εικόνα 3.16 : (α)-(γ) : Κέντρα των πυρήνων που περιέχουν ακτινογραφίες ασθενών με πνευμονία, (δ) : Κέντρο πυρήνα που περιέχει ακτινογραφίες υγιών ανθρώπων, (ε) : Κέντρο πυρήνα που περιέχει ακτινογραφίες ασθενών με COVID-19

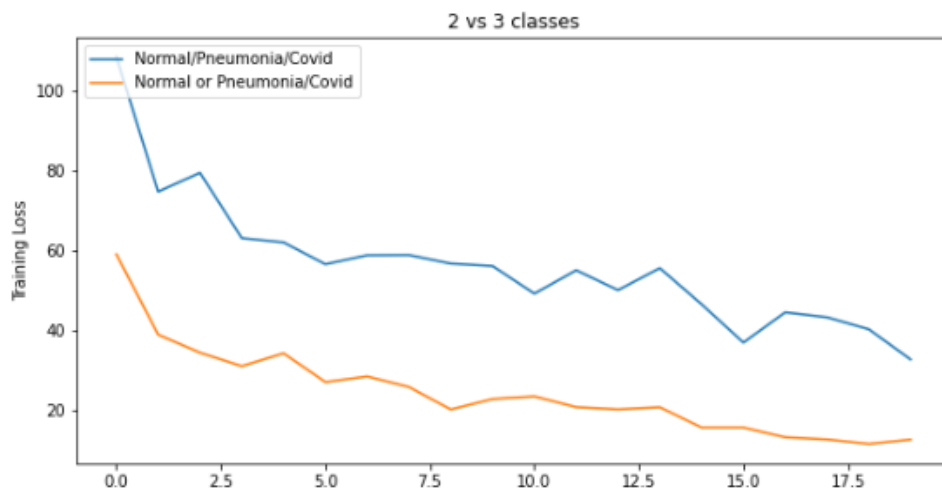
Όπως φαίνεται και στην παραπάνω εικόνα προκύπτουν τρεις πυρήνες για ασθενείς με πνευμονία (3.16.α, β, γ), ένας για υγιείς ανθρώπους (3.16.δ) και ένας για ασθενείς που πάσχουν από COVID-19 (3.16.ε).

### 3.7 Κατηγοριοποίηση σε 2 Κλάσεις

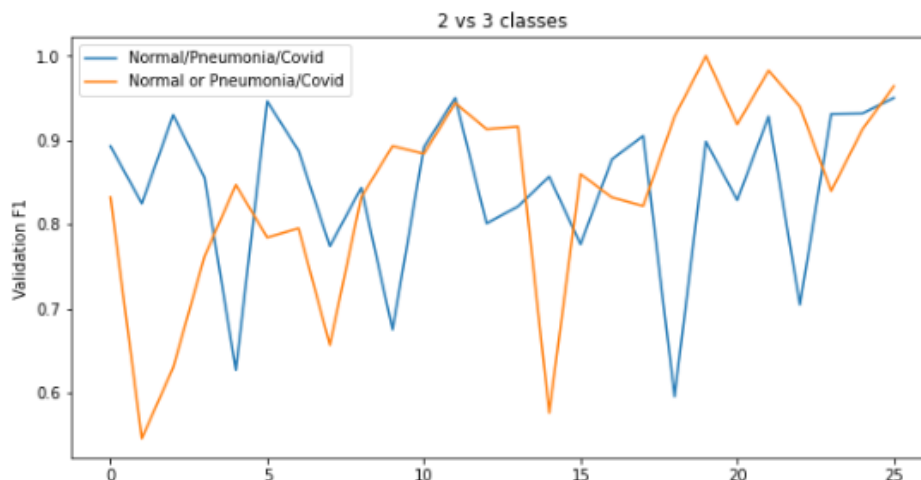
Ένα επιπλέον πείραμα που θα είχε ενδιαφέρον να εκτελέσουμε θα ήταν να εξετάσουμε την απόδοση του δικτύου όταν τα δεδομένα κατηγοριοποιούνται σε 2 κλάσεις. Αυτές οι δύο κλάσεις θα είναι ασθενείς που πάσχουν από COVID-19 και ασθενείς που δεν πάσχουν από COVID-19. Συγχωνεύουμε δηλαδή τα δεδομένα των κλάσεων υγιείς και πνευμονία των προηγούμενων κεφαλαίων στην κλάση των μη-πασχόντων από COVID-19.

Το μοντέλο που χρησιμοποιούσαμε πριν τροποποιείται έτσι ώστε να έχει μόνο δύο εξόδους στο τελευταίο επίπεδο.

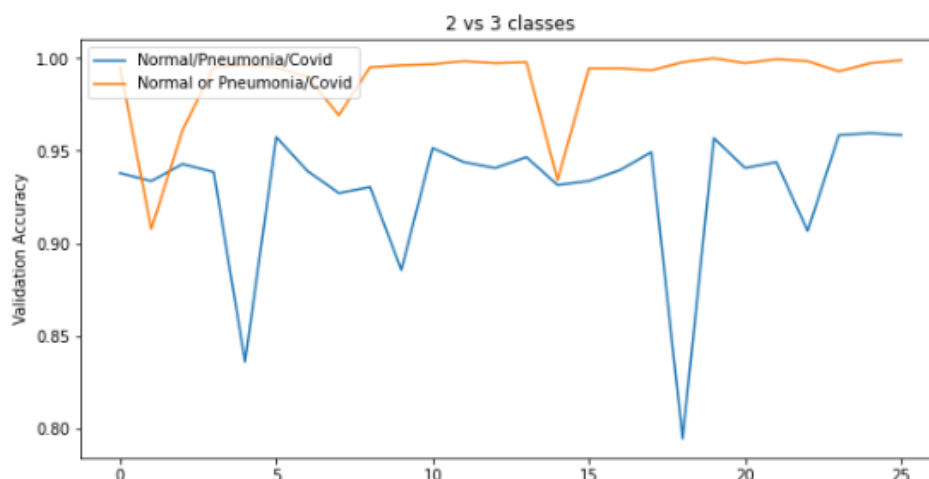
Παρακάτω παρουσιάζουμε τα διαγράμματα που προκύπτουν κατά την εκπαίδευση του δικτύου για τις μετρικές training loss και F1 score και accuracy στο validation set σε σύγκριση με ένα μοντέλο που κατηγοριοποιεί σε 3 κλάσεις.



Εικόνα 3.17 : Training loss μοντέλων που κατηγοριοποιούν σε 2 ή 3 κλάσεις



Εικόνα 3.18 : Validation F1 μοντέλων που κατηγοριοποιούν σε 2 ή 3 κλάσεις



Εικόνα 3.18 : Validation Accuracy μοντέλων που κατηγοριοποιούν σε 2 ή 3 κλάσεις

Όπως παρατηρούμε, αλλά και όπως ήταν αναμενόμενο, όταν το μοντέλο αντιμετωπίζει το ευκολότερο πρόβλημα της κατηγοριοποίησης σε δύο μόνο κλάσεις αποδίδει καλύτερα. Κατά την 17η εποχή παρατηρούμε ότι το μοντέλο πετυχαίνει f1 score ίσο με 1 στα δεδομένα του validation set.

### 3.8 Συσταδοποίηση σε 2 Κλάσεις

Έχοντας στη διάθεσή μας και ένα μοντέλο που ταξινομεί τα δεδομένα σε δύο κλάσεις, θα επιχειρήσουμε να εφαρμόσουμε και πάλι clustering, ακολουθώντας παρόμοια διαδικασία με πριν.

Εφαρμόζοντας k-means clustering προκύπτει ότι ο ιδανικός αριθμός πυρήνων k είναι και εδώ 5. Παρακάτω, παρουσιάζουμε τα αποτελέσματα που προκύπτουν όταν χρησιμοποιούμε clustering για να κάνουμε προβλέψεις για τα δεδομένα μας. Τα δεδομένα παρουσιάζονται σε μορφή classification reports.

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.82	0.93	0.87	127
Δεν πάσχει από COVID - 19	1.00	1.00	1.00	16595
accuracy			1.00	16722
Macro average	0.91	0.96	0.93	16722
Weighted average	1.00	1.00	1.00	16722

Εικόνα 3.19 : Classification Report μοντέλου εκπαιδευμένου σε ακτινογραφίες για την κατηγοριοποίηση σε 2 κλάσεις στο train set του X-ray dataset, χρησιμοποιώντας clustering

	precision	recall	F1-score	support
Πάσχει από COVID -19	1.00	1.00	1.00	14
Δεν πάσχει από COVID - 19	1.00	1.00	1.00	1840
Accuracy			1.00	1854
Macro average	1.00	1.00	1.00	1854
Weighted average	1.00	1.00	1.00	1854

Εικόνα 3.20 : Classification Report μοντέλου εκπαιδευμένου σε ακτινογραφίες για την κατηγοριοποίηση σε 2 κλάσεις στο validation set του X-ray dataset, χρησιμοποιώντας clustering

Παρατηρούμε ότι σε σχέση με το μοντέλο για τις 3 κλάσεις η χρήση του μοντέλου για τις 2, πετυχαίνει λίγο χειρότερα αποτελέσματα στο train set αλλά καλύτερα αποτελέσματα στο validation set.

Στο επόμενο κεφάλαιο, θα εξετάσουμε πειράματα που αφορούν σύνολα δεδομένων που περιέχουν αξονικές τομογραφίες.

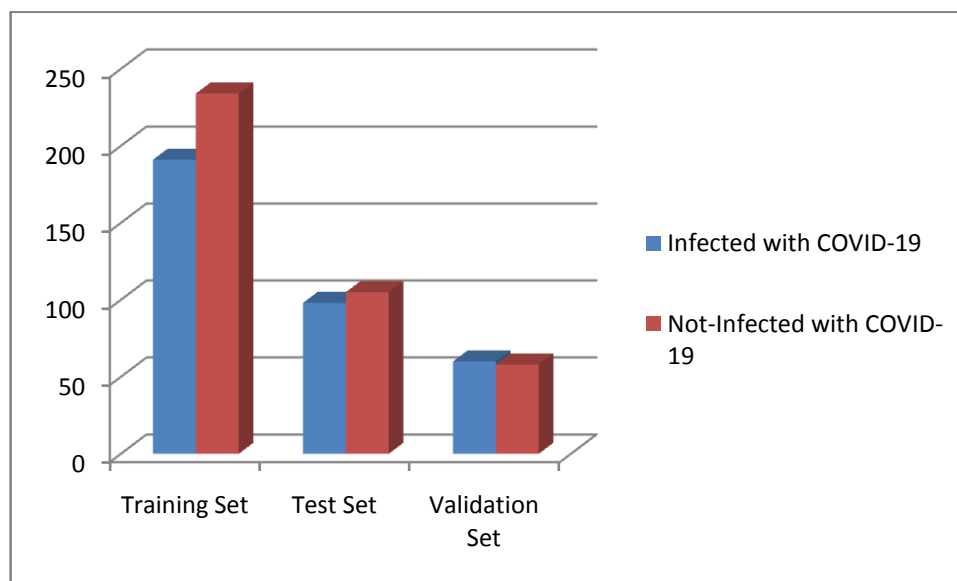
## 4. Σύνολο Δεδομένων με Αξονικές Τομογραφίες

### 4.1 Αρχιτεκτονική

Στο κεφάλαιο αυτό εξετάζουμε την απόδοση του μοντέλου μας σε ένα σύνολο δεδομένων που περιέχει αξονικές τομογραφίες. Καθώς οι τομογραφίες του συνόλου αυτού είναι επισημασμένες με 2 κλάσεις ( ασθενείς που πάσχουν ή όχι από COVID-19) η αρχιτεκτονική του δικτύου που θα χρησιμοποιήσουμε είναι ίδια με αυτήν του βήματος 3.7 του προηγούμενου κεφαλαίου.

### 4.2 COVID-CT Dataset

Το σύνολο δεδομένων που θα εξετάσουμε αποτελείται από τoμές αξονικών τομογραφιών και περιέχει 349 αξονικές τομογραφίες από 216 ασθενείς που εξετάστηκαν θετικοί στον COVID-19 και 397 αξονικές τομογραφίες ανθρώπων που δεν πάσχουν από COVID-19 και είναι διαθέσιμο εδώ ( [32] ). Καθώς η κατανομή των κλάσεων στο dataset είναι αρκετά ομοιόμορφη θα χρησιμοποιήσουμε training, test και validation sets. Η κατανομή των δεδομένων σε αυτά φαίνεται παρακάτω.



Εικόνα 4.1 : Κατανομές training, validation και test sets

Συγκριμένα έχουμε

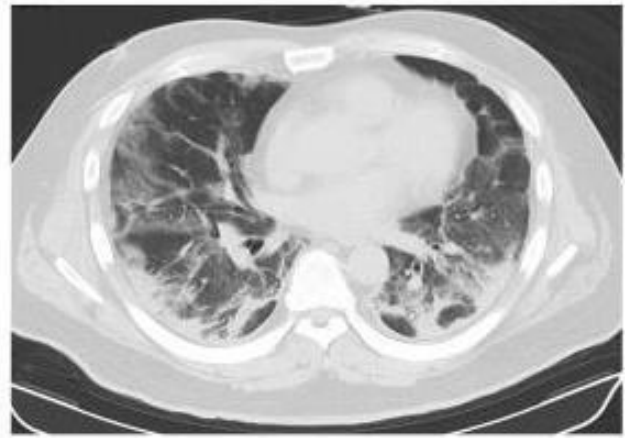
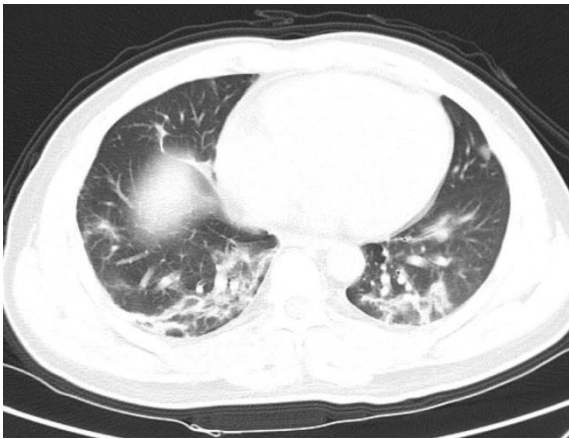
- στο training set 191 αξονικές τομογραφίες ανθρώπων που πάσχουν από COVID-19 και 234 αξονικές τομογραφίες ανθρώπων που δεν πάσχουν από COVID-19
- στο test set 98 αξονικές τομογραφίες ανθρώπων που πάσχουν από COVID-19 και 105 αξονικές τομογραφίες ανθρώπων που δεν πάσχουν από COVID-19



- στο validation set 60 αξονικές τομογραφίες ανθρώπων που πάσχουν από COVID-19 και 58 αξονικές τομογραφίες ανθρώπων που δεν πάσχουν από COVID-19

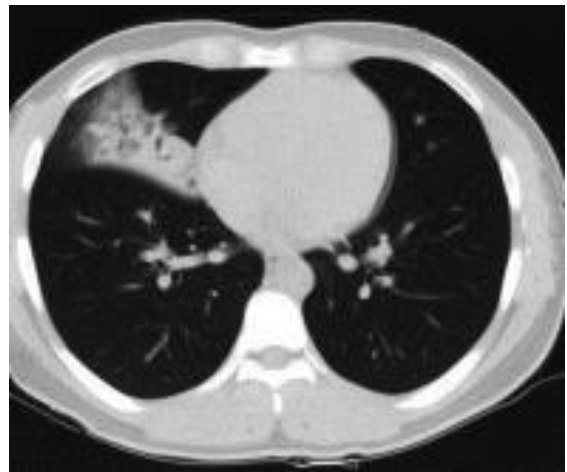
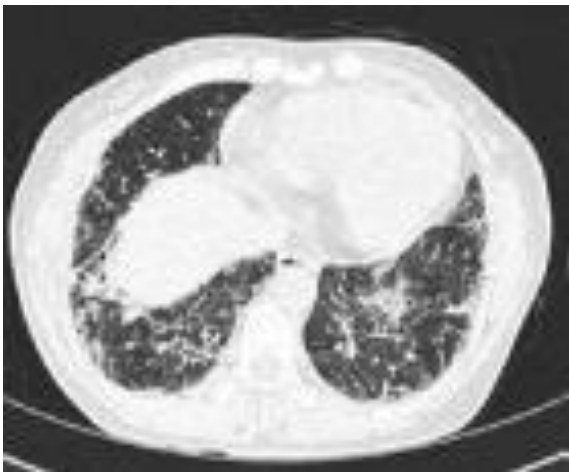
Παρακάτω παρουσιάζουμε ενδεικτικά κάποιες αξονικές τομογραφίες από την κάθε κλάση.

- Πάσχουν από COVID-19



Εικόνα 4.2 : Αξονικές τομογραφίες ασθενών με COVID-19

- Δεν πάσχουν από COVID-19



Εικόνα 4.3 : Αξονικές τομογραφίες ανθρώπων αρνητικών στον COVID-19

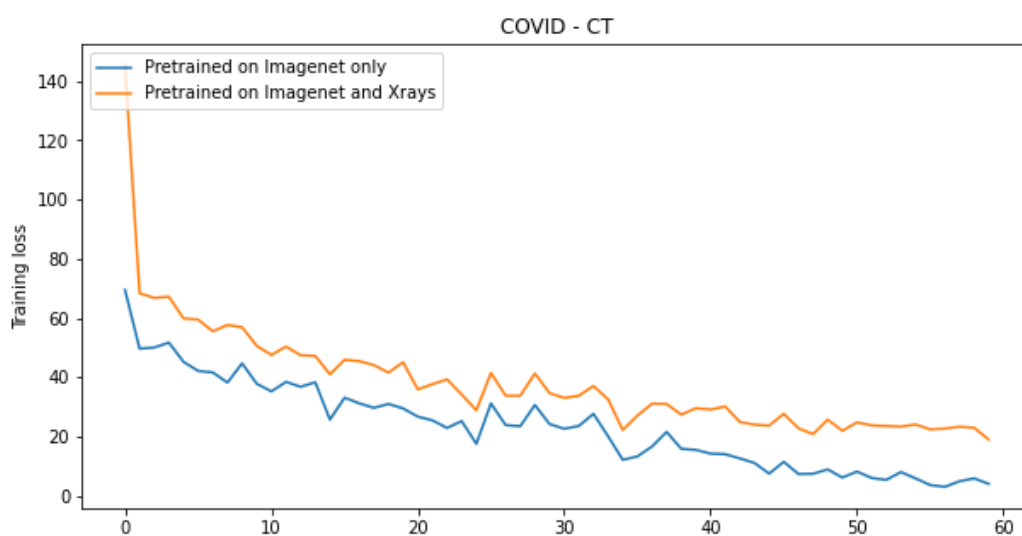
Όπως και πριν στις ακτινογραφίες, έτσι και εδώ επειδή οι αξονικές τομογραφίες έχουν μεγάλο μέγεθος για να τις δώσουμε ως είσοδο στο νευρωνικό δίκτυο τις μετασηματίσαμε σε διαστάσεις 256x256.

Ακόμη, για να αποφύγουμε το overtraining κατά τη διάρκεια της εκπαίδευσης εφαρμόζουμε παρόμοιους μετασηματισμούς με αυτούς που εφαρμόσαμε προηγουμένως.

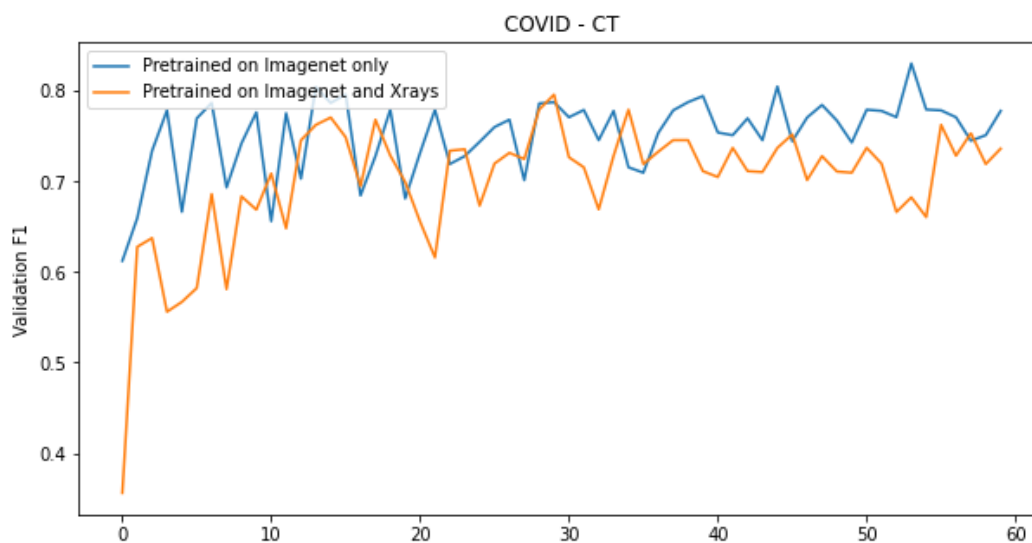
### 4.3 Χρήση Προ-εκπαιδευμένων ή μη Δικτύων (Pretrained vs Not-Pretrained)

Όπως είπαμε και στην αρχή τα μοντέλα που χρησιμοποιούμε είναι pretrained στο Imagenet. Έχοντας όμως διαθέσιμα και δεδομένα εκπαίδευσης ακτινογραφιών έχει ενδιαφέρον να δούμε αν έχει νόημα πέρα από την εκπαίδευση στο Imagenet, να εκπαιδεύσουμε τα μοντέλα μας και στις ακτινογραφίες πριν προχωρήσουμε σε εκπαίδευση στις αξονικές τομογραφίες.

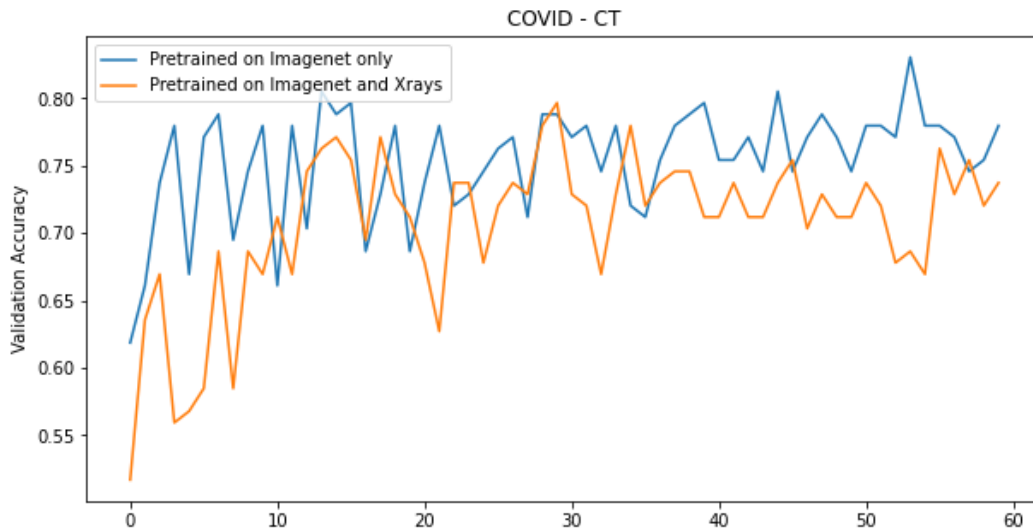
Παρακάτω παρουσιάζουμε τα αποτελέσματα της σύγκρισης δύο μοντέλων, ενός εκπαιδευμένου μόνο στο Imagenet και ενός εκπαιδευμένου στο Imagenet και το Chest X-ray Dataset που εξετάσαμε πριν.



Εικόνα 4.4 : Training loss



Εικόνα 4.5 : F1 score στο validation set



Εικόνα 4.6 : Accuracy score στο validation set

Παρατηρούμε ότι στο validation set το δίκτυο που έχει γίνει pretrained μόνο στο Imagenet αποδίδει λίγο καλύτερα.

Στη συνέχεια εξετάζουμε την απόδοση των δύο παραπάνω μοντέλων στο test set.

Αρχικά, παρουσιάζουμε το classification report του μοντέλου που είναι εκπαιδευμένο μόνο στο Imagenet.

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.85	0.72	0.78	98
Δεν πάσχει από COVID - 19	0.77	0.88	0.82	105
accuracy			0.80	203
Macro average	0.81	0.80	0.80	203
Weighted average	0.81	0.80	0.80	203

Εικόνα 4.7 : Classification Report μοντέλου εκπαιδευμένου στο Imagenet

Έπειτα, το classification report του μοντέλου που είναι εκπαιδευμένο αρχικά στο Imagenet και έπειτα στο Chest X-ray Dataset.

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.84	0.77	0.80	98
Δεν πάσχει από COVID - 19	0.80	0.87	0.83	105
accuracy			0.82	203
Macro average	0.82	0.82	0.82	203
Weighted average	0.82	0.82	0.82	203

Εικόνα 4.7 : Classification Report μοντέλου εκπαιδευμένου στο Imagenet και στο Chest X-ray dataset

Παρατηρούμε ότι το μοντέλο που είναι εκπαιδευμένο στο Imagenet και στο Chest X-ray Dataset πετυχαίνει λίγο καλύτερη απόδοση στο test set. Συγκεκριμένα, πετυχαίνει

F1 0.82 έναντι του μοντέλου που είναι εκπαιδευμένο μόνο στο Imagenet και πετυχαίνει F1 0.80.

Στη δημοσίευση Sample-Efficient Deep Learning for COVID-19 Diagnosis Based on CT Scans που είναι διαθέσιμη εδώ ( [33] ) από την ομάδα που είναι υπεύθυνη για την συλλογή των δεδομένων αναφέρουν ότι χρησιμοποιώντας ένα δίκτυο ResNet-50 pretrained στο Imagenet πετυχαίνουν 0.81 F1 και 0.80 accuracy, αποτελέσματα πολύ παρόμοια με αυτά που πετύχαμε και εμείς.

#### 4.4 Συσταδοποίηση στην Ανάλυση Αξονικών Τομογραφιών

Θα είχε ενδιαφέρον να εφαρμόσουμε την τεχνική του clustering και στο σύνολο δεδομένων με τις αξονικές τομογραφίες. Όπως και πριν για να πετύχουμε το clustering θα περάσουμε τα δεδομένα εκπαίδευσης από το μοντέλο που έχουμε εκπαιδεύσει αφαιρώντας το τελευταίο επίπεδο και κρατώντας τις εξόδους που θα προκύψουν. Στη συνέχεια, εφαρμόζουμε k-means clustering. Με τον τρόπο αυτό, προκύπτει ότι για τις αξονικές τομογραφίες ο ιδανικός αριθμός πυρήνων είναι 3. Προκύπτουν 2 πυρήνες που αντιστοιχούν σε αξονικές τομογραφίες ανθρώπων χωρίς COVID-19 και 1 πυρήνας για ασθενείς με COVID-19.

Παρουσιάζουμε παρακάτω τα αποτελέσματα από την εφαρμογή clustering για να πάρουμε τις προβλέψεις στα train και validation sets του συνόλου δεδομένων με τις αξονικές τομογραφίες.

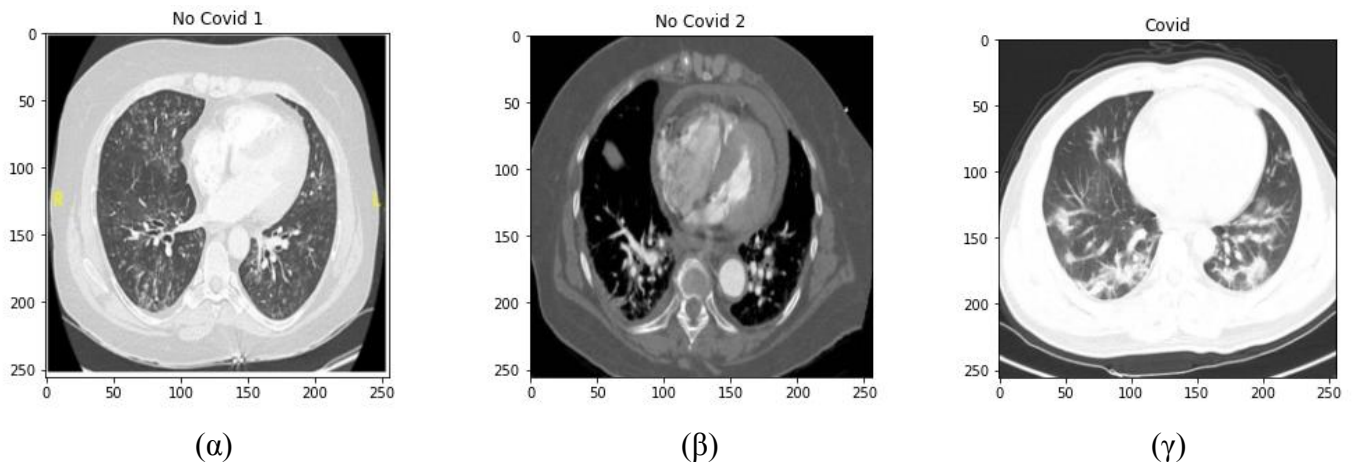
	precision	recall	F1-score	support
Πάσχει από COVID -19	1.00	1.00	1.00	191
Δεν πάσχει από COVID - 19	1.00	1.00	1.00	234
accuracy			1.00	425
Macro average	1.00	1.00	1.00	425
Weighted average	1.00	1.00	1.00	425

Εικόνα 4.8 : Classification Report μοντέλου εκπαιδευμένου σε αξονικές τομογραφίες στο train set του COVID-CT dataset, χρησιμοποιώντας clustering

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.84	0.82	0.83	60
Δεν πάσχει από COVID - 19	0.82	0.84	0.83	58
Accuracy			0.83	118
Macro average	0.83	0.83	0.83	118
Weighted average	0.83	0.83	0.83	118

Εικόνα 4.9 : Classification Report μοντέλου εκπαιδευμένου σε αξονικές τομογραφίες στο validation set του COVID-Ct dataset, χρησιμοποιώντας clustering

Παρατηρούμε ότι χρησιμοποιώντας clustering πετυχαίνουμε πολύ καλά αποτελέσματα για την κατηγοριοποίηση των δεδομένων εκπαίδευσης, ενώ στο validation set πετυχαίνουμε f1 score 0.83, ίδιο με αυτό που πετυχαίνουμε χωρίς την εφαρμογή του clustering, έχοντας όμως τα επιπρόσθετα οφέλη όπως η οπτικοποίηση των πυρήνων που βλέπουμε στη συνέχεια.



Εικόνα 4.10 : (α),(β) : Κέντρα των πυρήνων που περιέχουν αξονικές τομογραφίες ασθενών αρνητικών στον κορονοϊό, (γ) : Κέντρο πυρήνα που περιέχει αξονικές τομογραφίες ασθενών θετικών στον κορονοϊό

Στο επόμενο κεφάλαιο θα εξετάσουμε την απόδοση των μοντέλων μας σε σύνολο δεδομένων που περιέχει ακτινογραφίες αλλά και αξονικές τομογραφίες.

## 5. Σύνολο Δεδομένων με Ακτινογραφίες και Αξονικές Τομογραφίες

### 5.1 Αρχιτεκτονική

Η αρχιτεκτονική του μοντέλου που θα χρησιμοποιήσουμε και για αυτά τα πειράματα είναι αυτή με το επιπλέον επίπεδο των 32 νευρώνων και τις 2 εξόδους στο τελευταίο επίπεδο.

### 5.2 Dataset

Το dataset που χρησιμοποιούμε προκύπτει από την συνένωση των δύο προηγούμενων. Καθώς το σύνολο δεδομένων με τις ακτινογραφίες χρησιμοποιεί 3 κλάσεις ενώ το σύνολο δεδομένων των αξονικών τομογραφιών 2, θεωρούμε ότι οι εικόνες που ανήκουν στις κλάσεις πνευμονία και υγιείς ανήκουν στην ίδια κλάση, την "δεν πάσχει από COVID-19", εφαρμόζοντας την ίδια λογική που χρησιμοποιήσαμε και στο κεφάλαιο 3.7.

### 5.3 Χρήση Μόνο Ακτινογραφιών ή Ακτινογραφιών & Αξονικών Τομογραφιών

Αρχικά, συγκρίνουμε ένα μοντέλο που έχει εκπαιδευτεί μόνο πάνω σε ακτινογραφίες με ένα μοντέλο που έχει εκπαιδευτεί σε ακτινογραφίες και αξονικές τομογραφίες. Για να πραγματοποιήσουμε την σύγκριση, θα εξετάσουμε την απόδοση των δύο μοντέλων στο Chest X-Ray dataset. Καθώς, για το σύνολο δεδομένων αυτό δεν χρησιμοποιήσαμε test set λόγω έλλειψης δεδομένων COVID-19, η σύγκριση θα γίνει χρησιμοποιώντας τα δεδομένα του validation set.

Παρακάτω παρουσιάζουμε τα classification reports των δύο μοντέλων.

Αρχικά, το classification report του μοντέλου εκπαιδευμένου σε ακτινογραφίες.

	precision	recall	F1-score	support
Πάσχει από COVID -19	1.00	1.00	1.00	1840
Δεν πάσχει από COVID - 19	1.00	1.00	1.00	14
accuracy			1.00	1854
Macro average	1.00	1.00	1.00	1854
Weighted average	1.00	1.00	1.00	1854

Εικόνα 5.1 : Classification Report μοντέλου εκπαιδευμένου σε ακτινογραφίες

Όπως διαπιστώσαμε και στο κεφάλαιο 3.7 η απόδοση ενός μοντέλου που εκπαιδύεται σε ακτινογραφίες που ανήκουν μόνο σε δύο κλάσεις είναι πάρα πολύ καλή, καθώς πετυχαίνει F1 score ίσο με 1.

Στη συνέχεια ακολουθεί το classification report του μοντέλου που εκπαιδύεται στο συνενωμένο σύνολο δεδομένων.

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.88	0.50	0.64	14
Δεν πάσχει από COVID - 19	1.00	1.00	1.00	1840
accuracy			1.00	1854
Macro average	0.94	0.75	0.82	1854
Weighted average	1.00	1.00	1.00	1854

Εικόνα 5.2 : Classification Report μοντέλου εκπαιδευμένου σε ακτινογραφίες και αξονικές τομογραφίες

Βλέπουμε ότι χρησιμοποιώντας το δεύτερο μοντέλο η απόδοση πέφτει δραματικά, καθώς από F1 score 1.00 για την κλάση COVID-19 η μετρική πέφτει στο 0.64. Κάτι τέτοιο, είναι βέβαιο αναμενόμενο αφού πλέον υπάρχει μεγάλη ανισοκατανομή τόσο ανάμεσα στις κλάσεις (COVID-19 ή όχι COVID-19) όσο και στον αριθμό των δειγμάτων που έχει το κάθε σύνολο δεδομένων και προς το παρών δεν εφαρμόζουμε κάποια τεχνική για να αντιμετωπίσουμε το πρόβλημα αυτό.

#### 5.4 Σύγκριση της Χρήσης μόνο Αξονικών Τομογραφιών ή και Ακτινογραφιών

Έπειτα θα συγκρίνουμε την απόδοση ενός μοντέλου που έχει εκπαιδευτεί μόνο σε αξονικές τομογραφίες με την απόδοση ενός μοντέλου που έχει εκπαιδευτεί σε ακτινογραφίες και αξονικές τομογραφίες. Για την σύγκριση θα χρησιμοποιήσουμε μετρικές υπολογισμένες πάνω στην απόδοση των δύο μοντέλων στο COVID-CT Dataset. Καθώς τα δεδομένα στην περίπτωση αυτή το επιτρέπουν θα εξετάσουμε την απόδοση των μοντέλων τόσο στο validation όσο και στο test set.

Αξίζει να σημειωθεί ότι για την σύγκριση χρησιμοποιούμε το μοντέλο που έχει εκπαιδευτεί μόνο στο Imagenet και στη συνέχεια στα CTs, και όχι αυτό που είναι pretrained αρχικά στο Imagenet και έπειτα στο Chest X-ray Dataset.

Παρακάτω παρουσιάζουμε τα classification reportstων δύο μοντέλων.

Αρχικά, τα classification reports του μοντέλου εκπαιδευμένου σε αξονικές τομογραφίες.

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.84	0.82	0.83	60
Δεν πάσχει από COVID - 19	0.82	0.84	0.83	58
accuracy			0.83	118
Macro average	0.83	0.83	0.83	118
Weighted average	0.83	0.83	0.83	118

Εικόνα 5.3 : Classification Report μοντέλου εκπαιδευμένου σε αξονικές τομογραφίες στο validation set του COVID-CT Dataset

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.85	0.72	0.78	98
Δεν πάσχει από COVID - 19	0.77	0.88	0.82	105
accuracy			0.80	203
Macro average	0.81	0.80	0.80	203
Weighted average	0.81	0.80	0.80	203

Εικόνα 5.4 : Classification Report μοντέλου εκπαιδευμένου σε αξονικές τομογραφίες στο test set του COVID-CT Dataset

Έπειτα, τα classification reports του μοντέλου εκπαιδευμένου σε αξονικές τομογραφίες και ακτινογραφίες.

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.60	0.95	0.74	60
Δεν πάσχει από COVID - 19	0.87	0.34	0.49	58
accuracy			0.65	118
Macro average	0.73	0.65	0.61	118
Weighted average	0.73	0.65	0.62	118

Εικόνα 5.5 : Classification Report μοντέλου εκπαιδευμένου σε αξονικές τομογραφίες και ακτινογραφίες στο validation set του COVID-CT Dataset

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.60	0.87	0.71	98
Δεν πάσχει από COVID - 19	0.79	0.46	0.58	105
accuracy			0.66	203
Macro average	0.69	0.66	0.64	203
Weighted average	0.70	0.66	0.64	203

Εικόνα 5.6 : Classification Report μοντέλου εκπαιδευμένου σε αξονικές τομογραφίες και ακτινογραφίες στο test set του COVID-CT Dataset

Όπως και στις ακτινογραφίες, έτσι και εδώ παρατηρούμε ότι η απόδοση του δικτύου πέφτει δραματικά όταν έχει εκπαιδευτεί σε ακτινογραφίας και αξονικές τομογραφίες. Εδώ, το πρόβλημα οφείλεται πιθανώς στο γεγονός ότι ο αριθμός των ακτινογραφιών είναι της τάξεως του 17 χιλιάδες ενώ ο αριθμός των αξονικών τομογραφιών είναι περίπου 750. Είναι λοιπόν, λογικό ένα μοντέλο που εκπαιδεύεται σε ένα σύνολο δεδομένων που περιέχει και τα δύο να μην αποδίδει καλά.

Στα επόμενα κεφάλαια θα εξετάσουμε τεχνικές που μπορούμε να εφαρμόσουμε για να αντιμετωπίσουμε αυτό το πρόβλημα, εστιάζοντας σε τρεις τεχνικές, το balancing, το alignment και το clustering.



## 5.5 Ισοροπημένες Κατηγορίες Δεδομένων

Μία πρώτη προσέγγιση που μπορούμε να εφαρμόσουμε για να αντιμετωπίσουμε τα προβλήματα που συναντήσαμε στις προηγούμενες ενότητες, είναι να εξισώσουμε τον αριθμό δειγμάτων που προέρχονται από την κάθε κλάση σε ένα batch εκπαίδευσης. Για να το πετύχουμε αυτό, θα πρέπει να εφαρμόσουμε υπερδειματοληψία στην κλάση με τα λιγότερα δείγματα. Στην περίπτωση μας, η κλάση αυτή είναι η κλάση των ασθενών που πάσχουν από COVID-19. Η υπερδειματοληψία εδώ γίνεται ανεξάρτητα από τον τύπο της κάθε εικόνας, δηλαδή αν πρόκειται για ακτινογραφία η αξονική τομογραφία. Έτσι, τόσο οι ακτινογραφίες όσο και οι αξονικές τομογραφίες ασθενών με COVID-19 παρουσιάζονται στο μοντέλο αρκετά περισσότερες από 1 φορές.

Αν θυμηθούμε την κατανομή των δειγμάτων στο COVID-CT Dataset, θυμόμαστε ότι περιέχει 349 αξονικές τομογραφίες ασθενών με COVID-19 και 349 τομογραφίες ανθρώπων χωρίς. Οι κλάσεις στο σύνολο δεδομένων αυτό, λοιπόν έχουν περίπου ίσα δείγματα. Καταλαβαίνουμε λοιπόν, ότι μία δειματοληψία όπως η παραπάνω μπορεί να διορθώσει το πρόβλημα της ανισοκατανομής δειγμάτων σε κλάσεις στο συνενωμένο dataset, αλλά θα "χαλάσει" ουσιαστικά την κατανομή των δειγμάτων σε κλάσεις στο COVID-CT dataset.

Έχοντας τα προβλήματα αυτά κατά νου, προχωράμε στην παρουσίαση των αποτελεσμάτων για ένα μοντέλο που εκπαιδεύτηκε με τον παραπάνω τρόπο. Τα αποτελέσματα παρουσιάζονται με την μορφή classification reports για το validation set του Chest X-ray dataset και τα validation και test sets του COVID-CT dataset.

	precision	recall	F1-score	support
Πάσχει από COVID -19	1.00	0.14	0.25	14
Δεν πάσχει από COVID - 19	0.99	1.00	1.00	1840
accuracy			0.99	1854
Macro average	1.00	0.57	0.62	1854
Weighted average	0.99	0.99	1.00	1854

Εικόνα 5.7 : Classification Report μοντέλου εκπαιδευμένου σε ακτινογραφίες και αξονικές τομογραφίες χρησιμοποιώντας balancing στις κλάσεις, στο validation set του X-ray dataset

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.58	0.63	0.61	60
Δεν πάσχει από COVID - 19	0.58	0.53	0.56	58
Accuracy			0.58	118
Macro average	0.58	0.58	0.58	118
Weighted average	0.58	0.58	0.58	118

Εικόνα 5.8 : Classification Report μοντέλου εκπαιδευμένου σε ακτινογραφίες και αξονικές τομογραφίες χρησιμοποιώντας balancing στις κλάσεις, στο validation set του COVID-CT dataset

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.58	0.64	0.61	98
Δεν πάσχει από COVID - 19	0.63	0.56	0.59	105
<hr/>				
accuracy			0.60	203
Macro average	0.60	0.60	0.60	203
Weighted average	0.60	0.60	0.60	203

Εικόνα 5.9 : Classification Report μοντέλου εκπαιδευμένου σε ακτινογραφίες και αξονικές τομογραφίες χρησιμοποιώντας balancing στις κλάσεις, στο test set του COVID-CT dataset

Παρατηρούμε ότι εφαρμόζοντας το συγκεκριμένο τρόπο balancing, όχι μόνο δεν βελτιώνεται η απόδοση αλλά γίνεται χειρότερη και από την απόδοση του μοντέλου χωρίς την εφαρμογή balancing και στα τρία sets (validation ακτινογραφιών και validation και test sets αξονικών τομογραφιών).

## 5.6 Ισοροπημένα Είδη Δεδομένων

Η επόμενη προσέγγιση που θα δοκιμάσουμε είναι να εφαρμόσουμε balancing όχι όμως ως προς τις κλάσεις αλλά ως προς το είδος της κάθε εικόνας. Δηλαδή σε κάθε batch να φροντίσουμε έτσι ώστε να έχουμε ίσο αριθμό τομογραφιών και αξονικών τομογραφιών. Έτσι η υπερδειγματοληψία που θα γίνει αυτή τη φορά θα αφορά μόνο το σύνολο δεδομένων COVID-CT, αλλά δεν θα πειράζει την εσωτερική κατανομή των δειγμάτων του στις κλάσεις του.

Παρακάτω παρουσιάζουμε τα classification reports στα ίδια sets με πριν, ενός μοντέλου που εκπαιδεύτηκε με τον παραπάνω τρόπο.

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.55	0.43	0.48	14
Δεν πάσχει από COVID - 19	1.00	1.00	1.00	1840
<hr/>				
accuracy			0.99	1854
Macro average	0.77	0.71	0.74	1854
Weighted average	0.99	0.99	0.99	1854

Εικόνα 5.10 : Classification Report μοντέλου εκπαιδευμένου σε ακτινογραφίες και αξονικές τομογραφίες χρησιμοποιώντας balancing στο είδος των εικόνων, στο validation set του X-ray dataset

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.63	0.95	0.75	60
Δεν πάσχει από COVID - 19	0.89	0.41	0.56	58
<hr/>				
Accuracy			0.69	118
Macro average	0.76	0.68	0.66	118
Weighted average	0.76	0.69	0.66	118

Εικόνα 5.11 : Classification Report μοντέλου εκπαιδευμένου σε ακτινογραφίες και αξονικές τομογραφίες χρησιμοποιώντας balancing στο είδος των εικόνων, στο validation set του COVID-CT dataset

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.59	0.94	0.72	98
Δεν πάσχει από COVID - 19	0.87	0.39	0.54	105
<hr/>				
accuracy			0.66	203
Macro average	0.73	0.66	0.63	203
Weighted average	0.74	0.66	0.63	203

Εικόνα 5.12 : Classification Report μοντέλου εκπαιδευμένου σε ακτινογραφίες και αξονικές τομογραφίες χρησιμοποιώντας balancing στο είδος των εικόνων, στο test set του COVID-CT dataset

Παρατηρούμε τώρα ότι η απόδοση στο COVID-CT dataset είναι λίγο καλύτερη χρησιμοποιώντας balancing στο είδος των εικόνων, κάτι αναμενόμενο καθώς αυτό το είδος balancing δεν χαλάει την εσωτερική κατανομή των κλάσεων.

Στο Chest X-ray dataset η απόδοση ανέβηκε επίσης, αλλά παραμένει χαμηλότερη από την περίπτωση που δεν χρησιμοποιούσαμε καθόλου balancing.

## 5.7 Ισορροπημένα Είδη και Κατηγορίες Δεδομένων

Η τρίτη και τελευταία τεχνική balancing που θα δοκιμάσουμε είναι balancing τόσο ως προς τις κλάσεις όσο και ως προς τον αριθμό των δειγμάτων. Συγκεκριμένα σε κάθε batch, φροντίζουμε να υπάρχουν ίσες εικόνες και από τα δύο είδη ( αξονικές τομογραφίες και ακτινογραφίες ) καθώς και ίσες εικόνες από τις δύο κλάσεις ( ασθενών COVID-19 και μη-ασθενών COVID-19).

Με αυτό τον τρόπο εξασφαλίζουμε ότι η κλάση που έχει λίγα δείγματα δηλαδή, η COVID-19 θα υποστεί υπερδειγματοληψία, χωρίς να πειράζουμε όμως την κατανομή των δειγμάτων στο COVID-CT dataset, αφού εκεί δεν υπάρχει πρόβλημα. Έτσι, η υπερδειγματοληψία ως προς τις κλάσεις θα αφορά κυρίως την κλάση COVID-19 στο σύνολο δεδομένων με τις ακτινογραφίες και . Η υπερδειγματοληψία που θα γίνει για τα είδη των εικόνων θα αφορά μόνο το COVID-CT dataset. Έτσι, δεν θα έχουμε το πρόβλημα που συναντήσαμε στο κεφάλαιο 5.5, όπου χαλούσε η κατανομή των αξονικών τομογραφιών ενώ ταυτόχρονα, θα φροντίσουμε ώστε όλες οι κλάσεις να αντιπροσωπεύονται ισότιμα σε αντίθεση με την τεχνική που ακολουθήσαμε στο κεφάλαιο 5.6.

Παρακάτω παρουσιάζουμε τα classification reports με τα αποτελέσματα ενός μοντέλου που εκπαιδεύτηκε με την παραπάνω τεχνική.

	precision	recall	F1-score	support
Πάσχει από COVID -19	1.00	0.36	0.53	14
Δεν πάσχει από COVID - 19	1.00	1.00	1.00	1840
accuracy			1.00	1854
Macro average	1.00	0.68	0.76	1854
Weighted average	1.00	1.00	0.99	1854

Εικόνα 5.13 : Classification Report μοντέλου εκπαιδευμένου σε ακτινογραφίες και αξονικές τομογραφίες χρησιμοποιώντας balancing στο είδος των εικόνων και στις κλάσεις, στο validation set του X-ray dataset

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.74	0.93	0.82	60
Δεν πάσχει από COVID - 19	0.90	0.66	0.76	58
Accuracy			0.80	118
Macro average	0.82	0.79	0.79	118
Weighted average	0.82	0.80	0.79	118

Εικόνα 5.14 : Classification Report μοντέλου εκπαιδευμένου σε ακτινογραφίες και αξονικές τομογραφίες χρησιμοποιώντας balancing στο είδος των εικόνων και στις κλάσεις, στο validation set του COVID-CT dataset

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.71	0.78	0.74	98
Δεν πάσχει από COVID - 19	0.77	0.70	0.74	105
accuracy			0.74	203
Macro average	0.74	0.74	0.74	203
Weighted average	0.74	0.74	0.74	203

Εικόνα 5.15 : Classification Report μοντέλου εκπαιδευμένου σε ακτινογραφίες και αξονικές τομογραφίες χρησιμοποιώντας balancing στο είδος των εικόνων και στις κλάσεις, στο test set του COVID-CT dataset

Με τον τρόπο αυτό παρατηρούμε ότι βελτιώνεται το macro-averaged f1 score και στα τρία sets σε σχέση με τους προηγούμενους δύο τρόπους balancing. Κάτι τέτοιο ήταν αναμενόμενο καθώς ο τρόπος αυτός συνδυάζει τους προηγούμενους δύο αποφεύγοντας τα προβλήματα που δημιουργούνται αν εφαρμοσθούν μεμονωμένα.

Στο COVID-CT dataset η απόδοση βελτιώνεται τόσο, που πλησιάζει την απόδοση μοντέλων που έχουν εκπαιδευτεί μόνο στο dataset αυτό. Συγκεκριμένα, τότε είχαμε 0.83 και 0.80 f1 scores στα validation και test sets αντίστοιχα, ενώ τώρα πετυχαίνουμε 0.79 και 0.74. Η λίγο χαμηλότερη απόδοση είναι λογική αν αναλογιστεί κανείς ότι το μοντέλο αυτό έχει εκπαιδευτεί για να λύνει ουσιαστικά δύο προβλήματα ενώ το άλλο είναι πιο εξειδικευμένο.

Στο Chest X-ray dataset βλέπουμε ότι αν και η απόδοση του μοντέλου βελτιώνεται σε σχέση με τους προηγούμενους δύο τρόπους, παραμένει χαμηλότερη από το να μην

εφαρμόσουμε καθόλου balancing. Ενδεχομένως, το πρόβλημα αυτό να οφείλεται στο γεγονός ότι κατά την εκπαίδευση επαναλαμβάνεται κάθε φορά ένα δείγμα COVID-19, από τα οποία δεν υπάρχουν πολλά ( συγκεκριμένα 128 ), οπότε το μοντέλο καταλήγει να βλέπει πολλές φορές το ίδιο δείγμα και γίνεται overfit.

Στη συνέχεια, θα εξετάσουμε μια διαφορετική τεχνική για την εκπαίδευση του μοντέλου στο συνενωμένο dataset, το alignment.

## 5.8 Ευθυγράμμιση του Είδους των Δεδομένων

Μία άλλη τεχνική που μπορούμε να δοκιμάσουμε είναι αντί να αφήνουμε τα δείγματα σε κάθε batch να έρχονται τυχαία, να φροντίζουμε ώστε να υπάρχει μία σταθερή αναλογία. Δηλαδή αν είχαμε παραδείγματος χάρι 200 δείγματα από την κλάση A, 10 δείγματα από την κλάση B και batch size 21, θα είχαμε 20 δείγματα τύπου A και 1 δείγμα τύπου B σε κάθε batch. Συνολικά, θα χρειαζόμασταν 10 επαναλήψεις για να καλύψουμε το σύνολο των δεδομένων. Η τεχνική αυτή ονομάζεται alignment.

Αρχικά, θα εφαρμόσουμε alignment ως προς τις κλάσεις COVID-19 και όχι COVID-19, όπως κάναμε και πριν, ανεξάρτητα από το είδος της εικόνας. Θυμόμαστε ότι έχουμε 127 ακτινογραφίες και 191 αξονικές τομογραφίες από ασθενείς με COVID-19 στα δεδομένα εκπαίδευσης, δηλαδή συνολικά 318 εικόνες που αντιστοιχούν σε δείγματα COVID-19. Επίσης, έχουμε 16577 ακτινογραφίες και 234 αξονικές τομογραφίες, σύνολο 16811 εικόνες στα δεδομένα εκπαίδευσης που αντιστοιχούν σε ανθρώπους αρνητικούς στον COVID-19.

Για να πετύχουμε το alignment θα φροντίζουμε έτσι ώστε σε κάθε batch να διατηρείται σταθερή η αναλογία των κλάσεων, ώστε να είναι παρόντα δείγματα και από τις δύο κλάσεις.

Παρακάτω παρουσιάζουμε τα αποτελέσματα σε μορφή classification reports που προέκυψαν εφαρμόζοντας την τεχνική αυτή, στο validation set του X-ray dataset και στα validation και test sets του COVID-CT dataset.

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.58	0.79	0.67	14
Δεν πάσχει από COVID - 19	1.00	1.00	1.00	1840
<hr/>				
accuracy			0.99	1854
Macro average	0.79	0.89	0.83	1854
Weighted average	1.00	0.99	0.99	1854

Εικόνα 5.16 : Classification Report μοντέλου εκπαιδευμένου σε ακτινογραφίες και αξονικές τομογραφίες χρησιμοποιώντας alignment στις κλάσεις, στο validation set του X-ray dataset

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.64	0.90	0.74	60
Δεν πάσχει από COVID - 19	0.82	0.47	0.59	58
Accuracy			0.69	118
Macro average	0.73	0.68	0.67	118
Weighted average	0.73	0.69	0.67	118

Εικόνα 5.17 : Classification Report μοντέλου εκπαιδευμένου σε ακτινογραφίες και αξονικές τομογραφίες χρησιμοποιώντας alignment στις κλάσεις, στο validation set του COVID-CT dataset

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.60	0.70	0.65	98
Δεν πάσχει από COVID - 19	0.67	0.56	0.61	105
Accuracy			0.63	203
Macro average	0.64	0.63	0.63	203
Weighted average	0.64	0.62	0.63	203

Εικόνα 5.18 : Classification Report μοντέλου εκπαιδευμένου σε ακτινογραφίες και αξονικές τομογραφίες χρησιμοποιώντας alignment στις κλάσεις, στο test set του COVID-CT dataset

Παρατηρούμε ότι η εφαρμογή του alignment στις κλάσεις οδηγεί σε πολύ καλύτερα αποτελέσματα από ότι το αντίστοιχο balancing, καθώς δεν προκαλεί τις αρνητικές συνέπειες που προκαλεί το δεύτερο στην εσωτερική κατανομή των κλάσεων στο COVID-CT dataset.

Ιδιαίτερα, στο validation set του X-ray dataset παρατηρούμε ότι πετυχαίνει καλύτερη απόδοση και από τα 3 είδη balancing, καθώς και από την περίπτωση όπου δεν εφαρμόζουμε καμία από τις δύο τεχνικές. Στη συνέχεια θα εξετάσουμε δύο ακόμα τρόπους με τους οποίους μπορούμε να εφαρμόσουμε το alignment.

## 5.9 Ευθυγράμμιση των Κατηγοριών των Δεδομένων

Η επόμενη τεχνική που θα δοκιμάσουμε αφορά την εφαρμογή alignment, αλλά αυτή τη φορά ως προς το είδος των εικόνων. Αν έχουμε λοιπόν ένα παράδειγμα παρόμοιο με πριν, με δύο σύνολα δεδομένων, έστω A και B και 200 εικόνες από το σύνολο A και 10 εικόνες από το σύνολο B τότε φροντίζουμε σε κάθε batch των 21 εικόνων να έχουμε 20 εικόνες από το A και 1 εικόνα από το B.

Στην περίπτωση μας έχουμε το Chest X-ray dataset στο οποίο περιέχονται συνολικά 16704 ακτινογραφίες στα δεδομένα εκπαίδευσης και το COVID-CT dataset όπου συνολικά έχουμε 425 αξονικές τομογραφίες στα δεδομένα εκπαίδευσης. Με παρόμοιο τρόπο με το παραπάνω παράδειγμα, κρατάμε σε κάθε batch σταθερή την αναλογία δειγμάτων από το

κάθε dataset, ώστε να έχουμε κάθε φορά παρόντα δείγματα και από τα δύο σύνολα δεδομένων.

Παρακάτω παρουσιάζουμε τα αποτελέσματα που προκύπτουν από την εφαρμογή της παραπάνω τεχνικής, στο validation set του X-ray dataset και στα validation και test sets του COVID-CT dataset.

	precision	recall	F1-score	support
Πάσχει από COVID -19	1.00	0.64	0.78	14
Δεν πάσχει από COVID - 19	1.00	1.00	1.00	1840
accuracy			1.00	1854
Macro average	1.00	0.82	0.89	1854
Weighted average	1.00	1.00	1.00	1854

Εικόνα 5.19 : Classification Report μοντέλου εκπαιδευμένου σε ακτινογραφίες και αξονικές τομογραφίες χρησιμοποιώντας alignment στο είδος των εικόνων, στο validation set του X-ray dataset

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.62	0.92	0.74	60
Δεν πάσχει από COVID - 19	0.83	0.43	0.57	58
Accuracy			0.68	118
Macro average	0.73	0.67	0.66	118
Weighted average	0.73	0.68	0.66	118

Εικόνα 5.20 : Classification Report μοντέλου εκπαιδευμένου σε ακτινογραφίες και αξονικές τομογραφίες χρησιμοποιώντας alignment στο είδος των εικόνων, στο validation set του COVID-CT dataset

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.64	0.76	0.69	98
Δεν πάσχει από COVID - 19	0.73	0.61	0.66	105
Accuracy			0.68	203
Macro average	0.69	0.68	0.68	203
Weighted average	0.69	0.68	0.68	203

Εικόνα 5.21 : Classification Report μοντέλου εκπαιδευμένου σε ακτινογραφίες και αξονικές τομογραφίες χρησιμοποιώντας alignment στο είδος των εικόνων, στο test set του COVID-CT dataset

Εδώ παρατηρούμε πάλι μία βελτίωση στο validation set των ακτινογραφιών ενώ στις αξονικές τομογραφίες παρατηρούμε παρόμοια αλλά λίγο χειρότερη απόδοση από πριν. Στη συνέχεια, θα δοκιμάσουμε και τον τελευταίο τρόπο alignment, που αφορά alignment στις κλάσεις και στο είδος των εικόνων.

## 5.10 Ευθυγράμμιση των Κατηγοριών και του Είδους των Δεδομένων

Στο τρίτο και τελευταίο είδος alignment που θα εξετάσουμε, θα κρατήσουμε σε κάθε batch σταθερή την αναλογία δειγμάτων που προέρχονται από το κάθε dataset αλλά και της κάθε κλάσης. Για να γίνει πιο κατανοητό το παραπάνω θα δώσουμε και πάλι ένα παράδειγμα. Έστω ότι έχουμε δύο datasets A και B με 400 δείγματα το πρώτο και 20 δείγματα το δεύτερο. Επίσης, στο dataset A υπάρχουν 300 δείγματα που ανήκουν στην κλάση A1 και 100 δείγματα που ανήκουν στην κλάση B2. Στο dataset B υπάρχουν 10 δείγματα της κλάσης B1 και 10 δείγματα της κλάσης B2. Εφαρμόζοντας την παραπάνω τεχνική θα έχουμε σε κάθε batchτων 42 δειγμάτων , 40 δείγματα από το σύνολο δεδομένων A και 2 δείγματα από τα σύνολο δεδομένων B. Από τα 40 δείγματα του A τα 30 θα ανήκουν στην κλάση A1 και τα 10 στην κλάση B2, ενώ από το B θα έχουμε 1 δείγμα από την κλάση B1 και 1 δείγμα από την κλάση B2. Συνολικά, θα έχουμε 10 επαναλήψεις για να διατρέξουμε ολόκληρα τα δεδομένα εκπαίδευσης.

Στη συνέχεια θα δοκιμάσουμε την παραπάνω τεχνική στα δικά μας δεδομένα.

Εφαρμόζοντας την τεχνική αυτή στα δεδομένα του συνενωμένου dataset παίρνουμε τα αποτελέσματα που φαίνονται στα παρακάτω classification reports.

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.92	0.79	0.85	14
Δεν πάσχει από COVID - 19	1.00	1.00	1.00	1840
<hr/>				
accuracy			1.00	1854
Macro average	0.96	0.89	0.92	1854
Weighted average	1.00	1.00	1.00	1854

Εικόνα 5.19 : Classification Report μοντέλου εκπαιδευμένου σε ακτινογραφίες και αξονικές τομογραφίες χρησιμοποιώντας alignment στο είδος των εικόνων και στις κλάσεις, στο validation set του X-ray dataset

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.77	0.60	0.67	60
Δεν πάσχει από COVID - 19	0.66	0.81	0.73	58
<hr/>				
Accuracy			0.70	118
Macro average	0.71	0.71	0.70	118
Weighted average	0.71	0.70	0.70	118

Εικόνα 5.20 : Classification Report μοντέλου εκπαιδευμένου σε ακτινογραφίες και αξονικές τομογραφίες χρησιμοποιώντας alignment στο είδος των εικόνων και στις κλάσεις, στο validation set του COVID-CT dataset



	precision	recall	F1-score	support
Πάσχει από COVID -19	0.85	0.57	0.68	98
Δεν πάσχει από COVID - 19	0.69	0.90	0.79	105
Accuracy			0.74	203
Macro average	0.77	0.74	0.73	203
Weighted average	0.77	0.74	0.74	203

Εικόνα 5.21 : Classification Report μοντέλου εκπαιδευμένου σε ακτινογραφίες και αξονικές τομογραφίες χρησιμοποιώντας alignment στο είδος των εικόνων και στις κλάσεις, στο test set του COVID-CT dataset

Όπως και στο balancing έτσι και στο alignment, εφαρμόζοντας την τεχνική ως προς το είδος των εικόνων αλλά και ως προς τις κλάσεις παρατηρούμε ότι παίρνουμε τα καλύτερα αποτελέσματα.

Στη συνέχεια, πριν προχωρήσουμε στην εξαγωγή συμπερασμάτων, θα εξετάσουμε πως μπορούμε να χρησιμοποιήσουμε την τεχνική της ομαδοποίησης, για να δημιουργήσουμε μοντέλα που μπορούν να κατηγοριοποιήσουν εικόνες προερχόμενες και από τα δύο datasets.

## 5.11 Συσταδοποίηση

Η τελευταία τεχνική που θα εξετάσουμε για τη δημιουργία των μοντέλων που θέλουμε είναι το clustering. Όπως θυμόμαστε από τα κεφάλαια 3 και 4 έχουμε εξάγει κάποια κέντρα πυρήνων από ακτινογραφίες και αξονικές τομογραφίες που αντιστοιχούν σε ασθενείς με COVID-19 και σε ασθενείς χωρίς COVID-19. Μπορούμε να συνενώσουμε αυτούς τους πυρήνες και στη συνέχεια δίνοντας μία εικόνα στο σύστημα, που πρόκειται είτε για ακτινογραφία είτε για αξονική τομογραφία, να κάνουμε πρόβλεψη ανάλογα με την ετικέτα του κέντρου πυρήνα από το οποίο απέχει την ελάχιστη απόσταση.

Θυμόμαστε επίσης ότι για τις ακτινογραφίες, έχουμε χρησιμοποιήσει δύο μοντέλα για την εξαγωγή των πυρήνων, ένα που ταξινομεί τα δεδομένα σε 3 κλάσεις και ένα που τα ταξινομεί σε 2. Θα εξετάσουμε και τις δύο περιπτώσεις, χρησιμοποιώντας τους πυρήνες αυτούς εναλλάξ σε συνδυασμό με τους πυρήνες που προκύπτουν από τις αξονικές τομογραφίες. Επειδή και στις δύο περιπτώσεις είχαμε 5 πυρήνες που αντιστοιχούν σε ακτινογραφίες και ακόμη έχουμε 3 πυρήνες που αντιστοιχούν σε αξονικές τομογραφίες σε κάθε περίπτωση θα έχουμε  $5 + 3 = 8$  πυρήνες.

Παρακάτω παρουσιάζουμε τα αποτελέσματα που προκύπτουν από την εφαρμογή της παραπάνω τεχνικής σε μορφή classification reports για το validation set του Chest X-ray dataset και τα validation και test sets του COVID-CT dataset.

Αρχικά, τα αποτελέσματα που προκύπτουν χρησιμοποιώντας για πυρήνες ακτινογραφιών αυτούς που προκύπτουν από το μοντέλο με τις 2 εξόδους.

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.67	0.57	0.62	14
Δεν πάσχει από COVID - 19	1.00	1.00	1.00	1840
accuracy			0.99	1854
Macro average	0.83	0.78	0.81	1854
Weighted average	0.99	0.99	0.99	1854

Εικόνα 5.22 : Classification Report με clustering χρησιμοποιώντας για πυρήνες ακτινογραφιών αυτούς που προκύπτουν από το μοντέλο με τις 2 εξόδους, στο validation set του X-ray dataset

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.79	0.18	0.30	60
Δεν πάσχει από COVID - 19	0.53	0.95	0.68	58
Accuracy			0.56	118
Macro average	0.66	0.57	0.49	118
Weighted average	0.66	0.56	0.48	118

Εικόνα 5.23 : Classification Report με clustering χρησιμοποιώντας για πυρήνες ακτινογραφιών αυτούς που προκύπτουν από το μοντέλο με τις 2 εξόδους, στο validation set του COVID-CT dataset

	precision	recall	F1-score	support
Πάσχει από COVID -19	1.00	0.14	0.25	98
Δεν πάσχει από COVID - 19	0.56	1.00	0.71	105
Accuracy			0.59	203
Macro average	0.78	0.57	0.48	203
Weighted average	0.77	0.59	0.49	203

Εικόνα 5.24 : Classification Report με clustering χρησιμοποιώντας για πυρήνες ακτινογραφιών αυτούς που προκύπτουν από το μοντέλο με τις 2 εξόδους, στο validation set του COVID-CT dataset

Στη συνέχεια, τα αποτελέσματα που προκύπτουν χρησιμοποιώντας για πυρήνες ακτινογραφιών αυτούς από το μοντέλο με τις 3 εξόδους.

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.76	0.93	0.84	14
Δεν πάσχει από COVID - 19	1.00	1.00	1.00	1840
accuracy			1.00	1854
Macro average	0.88	0.96	0.92	1854
Weighted average	1.00	1.00	1.00	1854

Εικόνα 5.25 : Classification Report με clustering χρησιμοποιώντας για πυρήνες ακτινογραφιών αυτούς που προκύπτουν από το μοντέλο με τις 3 εξόδους, στο validation set του X-ray dataset

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.84	0.82	0.83	60
Δεν πάσχει από COVID - 19	0.82	0.84	0.83	58
Accuracy			0.83	118
Macro average	0.83	0.83	0.83	118
Weighted average	0.83	0.83	0.83	118

Εικόνα 5.26 : Classification Report με clustering χρησιμοποιώντας για πυρήνες ακτινογραφιών αυτούς που προκύπτουν από το μοντέλο με τις 3 εξόδους, στο validation set του COVID-CT dataset

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.85	0.72	0.78	98
Δεν πάσχει από COVID - 19	0.77	0.88	0.82	105
Accuracy			0.80	203
Macro average	0.81	0.80	0.80	203
Weighted average	0.81	0.80	0.80	203

Εικόνα 5.27 : Classification Report με clustering χρησιμοποιώντας για πυρήνες ακτινογραφιών αυτούς που προκύπτουν από το μοντέλο με τις 3 εξόδους, στο test set του COVID-CT dataset

Παρατηρούμε ότι, χρησιμοποιώντας τα κέντρα πυρήνων που προκύπτουν από το μοντέλο με τις 3 εξόδους τα αποτελέσματα είναι πολύ καλύτερα σε σχέση με το μοντέλο με τις 2 εξόδους. Η διαφορά γίνεται ιδιαίτερα εμφανής όταν κοιτάμε τα αποτελέσματα στο σύνολο δεδομένων με τις αξονικές τομογραφίες. Το γεγονός αυτό μπορεί να οφείλεται στο ότι τα κέντρα πυρήνων που προκύπτουν από το μοντέλο με τις εξόδους με τα κέντρα πυρήνων των αξονικών τομογραφιών μπορεί να είναι πολύ κοντά, οπότε όταν συνενώνονται ένα πιθανό δείγμα να "μπερδεύεται" και να ταξινομείται λάθος.

Στο επόμενο κεφάλαιο θα εξετάσουμε κάποια συμπεράσματα που προκύπτουν από την εφαρμογή και των 3 μεθόδων, balancing, alignment και clustering.

## 5.12 Συμπεράσματα

Τα συμπεράσματα που προκύπτουν από την εφαρμογή των 3 παραπάνω μεθόδων συνοψίζονται στα εξής :

- Τόσο στο balancing όσο και στο alignment τα καλύτερα αποτελέσματα τα έχουμε όταν τα εφαρμόζουμε και για τις κλάσεις αλλά και για το είδος των εικόνων.
- Με το alignment πετυχαίνουμε καλύτερα αποτελέσματα από το balancing για τα δεδομένα με τις ακτινογραφίες. Αυτό συμβαίνει καθώς με την τεχνική αυτή δεν

έχουμε υπερβολική υπερδειγματοληψία στην κλάση COVID-19 των ακτινογραφιών όπως έχουμε στο balancing.

- Με balancing σε είδη εικόνων και κλάσεις παίρνουμε καλύτερα αποτελέσματα από το αντίστοιχο alignment στις αξονικές τομογραφίες. Αυτό ίσως να συμβαίνει επειδή εφαρμόζοντας balancing τόσο σε είδη εικόνων όσο και σε κλάσεις δεν χαλάμε την εσωτερική κατανομή των κλάσεων στις αξονικές τομογραφίες, ενώ παράλληλα εξισώνουμε τον αριθμό τους με αυτόν των ακτινογραφιών.
- Τα καλύτερα αποτελέσματα και στα δύο datasets τα έχουμε όταν εφαρμόζουμε clustering χρησιμοποιώντας για πυρήνες ακτινογραφιών τους 5 πυρήνες του μοντέλου με τις 3 κλάσεις σε συνδυασμό με τους 3 πυρήνες των αξονικών τομογραφιών. Τότε πετυχαίνουμε f1 score 0.92 στο validation set των ακτινογραφιών και 0.83 και 0.80 f1 score στα validation και test sets των αξονικών τομογραφιών. Υπενθυμίζουμε ότι τα ίδια scores είχαμε στα αξονικές τομογραφίες χρησιμοποιώντας και μοντέλο εκπαιδευμένο μόνο σε αυτές.

Στο επόμενο κεφάλαιο θα δοκιμάσουμε κάποια πειράματα πάνω σε δεδομένα που πήραμε από ελληνικά νοσοκομεία.

## 6. Πειράματα σε ελληνικά δεδομένα

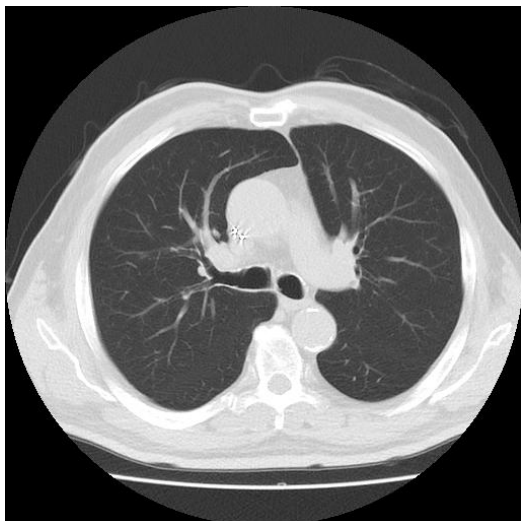
### 6.1 Αρχιτεκτονική

Καθώς τα δεδομένα που θα χρησιμοποιήσουμε είναι ταξινομημένα σε δύο κλάσεις, ασθενείς και μη-ασθενείς COVID-19 θα χρησιμοποιήσουμε το μοντέλο με τις 2 εξόδους και το επιπλέον επίπεδο με τα 32 layers. Το μοντέλο που θα χρησιμοποιήσουμε έχει εκπαιδευτεί στο Imagenet και στο COVID-CT dataset.

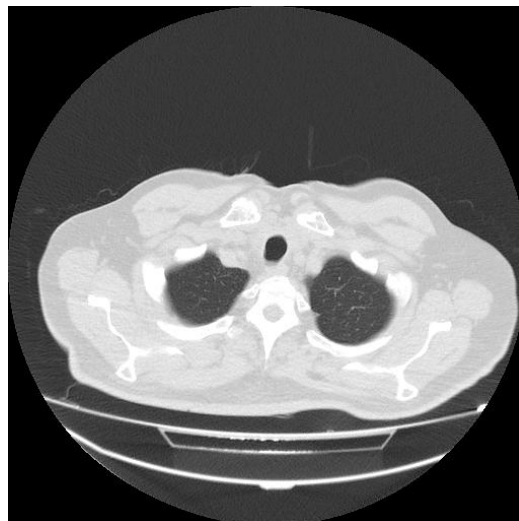
### 6.2 Σύνολο Δεδομένων

Το σύνολο δεδομένων που θα χρησιμοποιήσουμε περιέχει αξονικές τομογραφίες 12 ανθρώπων. Οι 6 από αυτούς έχουν διαγνωστεί θετικοί στον COVID-19 ενώ οι υπόλοιποι 6 όχι. Σε αντίθεση με το COVID-CT dataset όπου είχαμε μία τομή από την κάθε αξονική τομογραφία, εδώ έχουμε σειρά τομών που σχηματίζουν την αξονική τομογραφία ενός ασθενούς. Έτσι, για κάθε ασθενή έχουμε όχι μία, αλλά ίσως και εκατοντάδες εικόνες/τομές που αντιστοιχούν σε μία αξονική τομογραφία.

Παρακάτω παρουσιάζουμε ενδεικτικά δύο τομές από αξονικές τομογραφίες ενός ασθενούς με COVID-19 και ενός ανθρώπου αρνητικού στον COVID-19.



Εικόνα 6.1 : Τομή από αξονική τομογραφία ανθρώπου αρνητικό στον COVID-19



Εικόνα 6.2 : Τομή από αξονική τομογραφία ανθρώπου θετικό στον COVID-19

Τα δεδομένα τα χωρίζουμε σε train και test sets με το πρώτο να περιέχει 4 φακέλους ασθενών COVID-19 και 4 φακέλους μη-ασθενών COVID-19, και το δεύτερο 2 και 2 αντίστοιχα.

Στα επόμενα κεφάλαια θα εξετάσουμε κάποια πειράματα που κάναμε πάνω σε αυτά τα δεδομένα.

## 6.3 Χωρίς επανεκπαίδευση

Στο κεφάλαιο αυτό θα εξετάσουμε τεχνικές που εφαρμόσαμε χρησιμοποιώντας ένα μοντέλο εκπαιδευμένο στο COVID-CT dataset, κατευθειάν στα καινούρια δεδομένα χωρίς κάποια επανεκπαίδευση. Έτσι, ο διαχωρισμός σε train και test sets δεν έχει ιδιαίτερη σημασία στην περίπτωση αυτή.

Να αναφέρουμε επίσης ότι σε όλα τα πειράματα που θα ακολουθήσουν δεν αναφέρουμε τις μετρικές ανά εικόνα αλλά ανά φάκελο ασθενούς. Για να μετατρέψουμε τις προβλέψεις από προβλέψεις για κάθε εικόνα σε προβλέψεις για τον κάθε ασθενή εφαρμόζουμε 2 τεχνικές.

### 6.3.1 Με τεχνική "At least one"

Χρησιμοποιώντας την τεχνική αυτή, έχοντας τις προβλέψεις που αντιστοιχούν στην κάθε εικόνα, για να πάρουμε τις προβλέψεις για τον κάθε φάκελο ασθενούς, θεωρούμε ότι ο ασθενής έχει COVID-19 αν έστω και μία από τις εικόνες που του αντιστοιχούν έχει ετικέτα COVID-19, αλλιώς θεωρούμε ότι είναι αρνητικός.

Παρακάτω παρουσιάζουμε τα αποτελέσματα που προκύπτουν από την εφαρμογή της τεχνικής αυτής στα train και test sets σε μορφή classification reports.

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.67	1.00	0.80	4
Δεν πάσχει από COVID - 19	1.00	0.50	0.67	4
Accuracy			0.75	8
Macro average	0.83	0.75	0.73	8
Weighted average	0.83	0.75	0.73	8

Εικόνα 6.3 : Classification Report χωρίς επανεκπαίδευση με τεχνική "At least one" στο train set

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.50	1.00	0.67	2
Δεν πάσχει από COVID - 19	0.00	0.00	0.00	2
Accuracy			0.50	4
Macro average	0.25	0.50	0.33	4
Weighted average	0.25	0.50	0.33	4

Εικόνα 6.4 : Classification Report χωρίς επανεκπαίδευση με τεχνική "At least one" στο test set

Βλέπουμε ότι χρησιμοποιώντας την τεχνική αυτή στο train set παίρνουμε 0.73 f1 score που είναι ικανοποιητικό, αλλά στο test set δεν παίρνουμε καθόλου καλά αποτελέσματα καθώς όλες οι προβλέψεις είναι ασθενείς COVID-19, κάτι το οποίο δεν ισχύει. Έτσι, εκεί προκύπτει 0.33 f1 score.

### 6.3.2 Με τεχνική "Majority voting"

Χρησιμοποιώντας την τεχνική αυτή, αντί να προβλέπουμε COVID-19 όταν μία εικόνα ενός φακέλου έχει ετικέτα COVID-19, θα το κάνουμε μόνο όταν η πλειοψηφία των εικόνων έχουν COVID-19.

Έτσι, προκύπτουν τα ακόλουθα αποτελέσματα.

	precision	recall	F1-score	support
Πάσχει από COVID -19	1.00	1.00	1.00	4
Δεν πάσχει από COVID - 19	1.00	1.00	1.00	4
Accuracy			1.00	8
Macro average	1.00	1.00	1.00	8
Weighted average	1.00	1.00	1.00	8

Εικόνα 6.4 : Classification Report χωρίς επανεκπαίδευση με τεχνική "Majority Voting" στο train set

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.50	1.00	0.67	2
Δεν πάσχει από COVID - 19	0.00	0.00	0.00	2
Accuracy			0.50	4
Macro average	0.25	0.50	0.33	4
Weighted average	0.25	0.50	0.33	4

Εικόνα 6.5 : Classification Report χωρίς επανεκπαίδευση με τεχνική "Majority Voting" στο test set

Με την τεχνική αυτή, αν και παίρνουμε πολύ καλά αποτελέσματα στο train set, δηλαδή f1 score 1, στο test set το μοντέλο συνεχίζει να προβλέπει μόνο COVID-19 οπότε πάλι έχουμε 0.33 f1 score.

Καταλαβαίνουμε λοιπόν, ότι με το να εφαρμόζουμε το μοντέλο χωρίς κάποια επανεκπαίδευση δεν είναι πολύ αποτελεσματικό. Κάτι τέτοιο ήταν βέβαια αναμενόμενο, καθώς το μοντέλο είχε εκπαιδευτεί σε μεμονωμένες τομές αξονικών τομογραφιών ενώ εδώ πρόκειται για σειρές τομών.

## 6.4 Με επανεκπαίδευση

Στο κεφάλαιο αυτό θα χρησιμοποιήσουμε για τα πειράματα παρόμοιο μοντέλο με πριν, μόνο που τώρα θα το επανεκπαιδύσουμε στα δεδομένα εκπαίδευσης πριν το δοκιμάσουμε στο test set. Οι τεχνικές που θα δοκιμάσουμε για τις προβλέψεις ανά φάκελο είναι ίδιες με πριν.

#### 6.4.1 Με τεχνική "At least one"

Παρακάτω παρουσιάζουμε τα classification reports των μοντέλων που προκύπτουν χρησιμοποιώντας τεχνική "at least one" .

	precision	recall	F1-score	support
Πάσχει από COVID -19	1.00	1.00	1.00	4
Δεν πάσχει από COVID - 19	1.00	1.00	1.00	4
Accuracy			1.00	8
Macro average	1.00	1.00	1.00	8
Weighted average	1.00	1.00	1.00	8

Εικόνα 6.6 : Classification Report με επανεκπαίδευση με τεχνική "At least one" στο train set

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.50	1.00	0.67	2
Δεν πάσχει από COVID - 19	0.00	0.00	0.00	2
Accuracy			0.50	4
Macro average	0.25	0.50	0.33	4
Weighted average	0.25	0.50	0.33	4

Εικόνα 6.7 : Classification Report με επανεκπαίδευση με τεχνική "At least one" στο test set

Με επανεκπαίδευση και τεχνική "at least one" πετυχαίνουμε και πάλι άριστα αποτελέσματα στο train set, αλλά στο test set έχουμε f1 score 0.33 καθώς όλες οι προβλέψεις είναι COVID-19

#### 6.4.2 Με τεχνική "Majority voting"

Παρακάτω παρουσιάζουμε τα classification reports των μοντέλων που προκύπτουν χρησιμοποιώντας majority voting.

	precision	recall	F1-score	support
Πάσχει από COVID -19	1.00	1.00	1.00	4
Δεν πάσχει από COVID - 19	1.00	1.00	1.00	4
Accuracy			1.00	8
Macro average	1.00	1.00	1.00	8
Weighted average	1.00	1.00	1.00	8

Εικόνα 6.8 : Classification Report με επανεκπαίδευση με τεχνική "Majority Voting" στο train set



	precision	recall	F1-score	support
Πάσχει από COVID -19	0.67	1.00	0.80	2
Δεν πάσχει από COVID - 19	1.00	0.50	0.67	2
Accuracy			0.75	4
Macro average	0.83	0.75	0.73	4
Weighted average	0.83	0.75	0.73	4

Εικόνα 6.9 : Classification Report με επανεκπαίδευση με τεχνική "Majority Voting" στο test set

Παρατηρούμε ότι χρησιμοποιώντας majority voting και επανεκπαίδευση παίρνουμε τα καλύτερα αποτελέσματα έως τώρα αφού έχουμε άριστη ταξινόμηση στο train set ενώ πλέον στο test set δεν είναι όλες οι προβλέψεις COVID-19.

## 6.5 Συσταδοποίηση

Μία τελευταία τεχνική που θα είχε ενδιαφέρον να δοκιμάσουμε είναι clustering πάνω στα ελληνικά δεδομένα.

Για να το πετύχουμε αυτό θα χρησιμοποιήσουμε ένα δίκτυο που έχει επανεκπαιδευτεί στο train set των καινούριων δεδομένων, αλλά του έχουμε αφαιρέσει το τελευταίο layer. Έτσι, περνάμε μία ακόμη φορά τα δεδομένα εκπαίδευσης από το μοντέλο και κρατάμε τις εξόδους. Στη συνέχεια εφαρμόζουμε k-means clustering. Βρίσκουμε ότι ο ιδανικός αριθμός πυρήνας στην περίπτωση αυτή είναι 2, ένας για COVID-19 και ένας για αρνητικούς στον COVID-19.

Τα αποτελέσματα που παίρνουμε για k-means clustering με  $k = 2$  φαίνονται στα παρακάτω classification reports. Και εδώ όπως και πριν αναφέρουμε τις μετρικές για προβλέψεις ανά φάκελο και όχι ανά εικόνα. Καθώς προηγουμένως διαπιστώσαμε ότι η τεχνική που δίνει τα καλύτερα αποτελέσματα ήταν το majority voting αυτήν την τεχνική εφαρμόζουμε και στο clustering.

	precision	recall	F1-score	support
Πάσχει από COVID -19	1.00	1.00	1.00	4
Δεν πάσχει από COVID - 19	1.00	1.00	1.00	4
Accuracy			1.00	8
Macro average	1.00	1.00	1.00	8
Weighted average	1.00	1.00	1.00	8

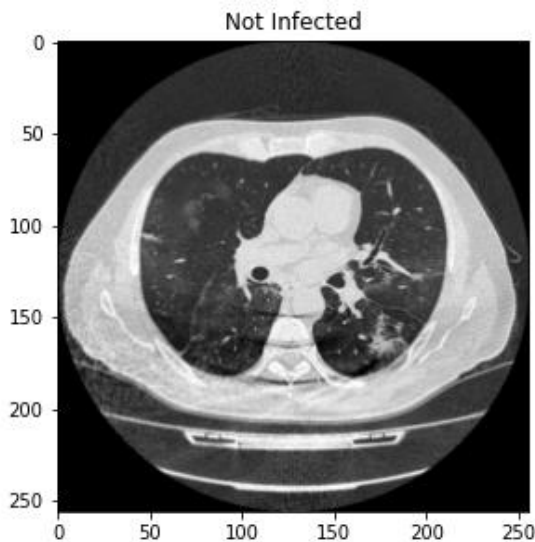
Εικόνα 6.10 : Classification Report με επανεκπαίδευση και clustering με τεχνική "Majority Voting" στο train set

	precision	recall	F1-score	support
Πάσχει από COVID -19	0.67	1.00	0.80	2
Δεν πάσχει από COVID - 19	1.00	0.50	0.67	2
Accuracy			0.75	4
Macro average	0.83	0.75	0.73	4
Weighted average	0.83	0.75	0.73	4

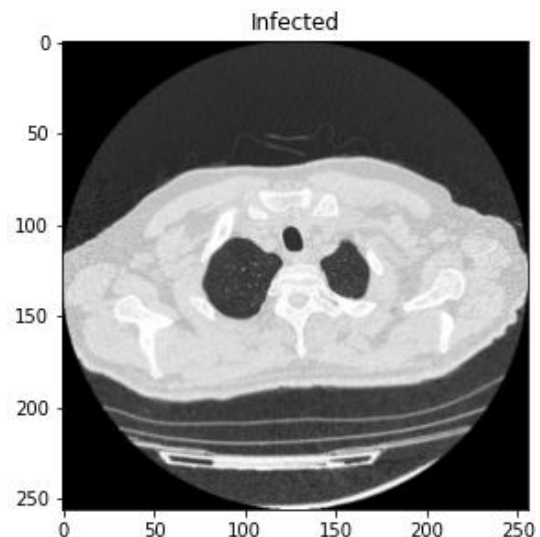
Εικόνα 6.11 : Classification Report με επανεκπαίδευση και clustering με τεχνική "Majority Voting" στο test set

Βλέπουμε ότι χρησιμοποιώντας clustering παίρνουμε τα ίδια αποτελέσματα που είχαμε εφαρμόζοντας επανεκπαίδευση και majority voting χωρίς clustering. Όπως έχουμε αναφέρει όμως η ομαδοποίηση μας προσφέρει και αρκετά επιπλέον αποτελέσματα.

Έτσι, μπορούμε να προχωρήσουμε στην οπτικοποίηση των κέντρων των πυρήνων που βρήκαμε πριν.



Εικόνα 6.12 : Κέντρο πυρήνα που περιέχει αξονικές τομογραφίες ανθρώπων αρνητικών στον COVID-19



Εικόνα 6.13 : Κέντρο πυρήνα που περιέχει αξονικές τομογραφίες ασθενών COVID-19

## 7. Σύνοψη

Στη διπλωματική αυτή προσπαθήσαμε να προτείνουμε νέους τρόπους για την διάγνωση του κορονοϊού.

Στο κεφάλαιο 2 κάναμε μία σύντομη εισαγωγή στις βασικές θεωρητικές έννοιες που χρειαστήκαμε. Συγκεκριμένα, παρουσιάσαμε τα συνελκτικά νευρωνικά δίκτυα που χρησιμοποιήσαμε στα πειράματά μας καθώς και την τεχνική της ομαδοποίησης την οποία εφαρμόσαμε στα δεδομένα μας.

Στο κεφάλαιο 3 παρουσιάσαμε αποτελέσματα πειραμάτων πάνω σε ένα σύνολο δεδομένων από ακτινογραφίες. Εξετάσαμε την απόδοση δικτύων που έχουν γίνει pretrained πάνω στο Imagenet τόσο στο πρόβλημα της κατηγοριοποίησης σε 2 αλλά και σε 3 κλάσεις. Εφαρμόσαμε επίσης clustering και για τις δύο περιπτώσεις και οπτικοποιήσαμε τα αποτελέσματά μας.

Στο κεφάλαιο 4 ακολουθήσαμε παρόμοια διαδικασία αυτήν τη φορά όμως για σύνολα δεδομένων που περιείχαν αξονικές τομογραφίες.

Στο κεφάλαιο 5 συνενώσαμε τα δύο προηγούμενα σύνολα δεδομένων σε ένα, και εξετάσαμε την απόδοση των δικτύων μας πάνω σε αυτό. Καθώς είδαμε ότι εκπαιδευοντας τα μοντέλα απλά πάνω στο συνενωμένο dataset χωρίς κάποια περαιτέρω τροποποίηση δεν πήραμε τα επιθυμητά αποτελέσματα δοκιμάσαμε διάφορες τεχνικές για να τα βελτιώσουμε. Συγκεκριμένα, εστίασαμε στις τεχνικές του balancing, του alignment αλλά και του clustering.

Στο κεφάλαιο 6 εξετάσαμε την απόδοση των μοντέλων μας πάνω σε ελληνικά δεδομένα. Εδώ δεν εξετάσαμε την κάθε εικόνα μεμονωμένα, αλλά φακέλους ασθενών με εκατοντάδες εικόνων. Για να πάρουμε προβλέψεις για αυτούς εφαρμόσαμε δύο τεχνικές, την "at least one" και την "majority voting". Τέλος, εφαρμόσαμε clustering και εδώ και οπτικοποιήσαμε τα κέντρα των πυρήνων.

Βασικό κίνητρο για την εκπόνηση της παρούσας διπλωματικής, αποτέλεσε ο ραγδαία αυξανόμενος αριθμός κρουσμάτων της πανδημίας του κορονοϊού και η επιτακτική ανάγκη για ανάπτυξη μεθοδολογιών γρήγορης και αποτελεσματικής διάγνωσης.

## Βιβλιογραφία

- [1] Y. Avrithis, N. Tsapatsoulis, and S. Kollias, "Broadcast news parsing using visual cues: A robust face detection approach," in *2000 IEEE International Conference on Multimedia and Expo, ICME 2000*, pp. 1469-1472.
- [2] K. Rapantzikos, N. Tsapatsoulis, Y. Avrithis, and S. Kollias, "Bottom-up spatiotemporal visual attention model for video analysis," *IET Image Processing*, vol. 1, no. 2, pp. 237-248.
- [3] D. Kollias, G. Marandianos, A. Raouzaiou, and A. Stafylopatis, "Interweaving deep learning and semantic techniques for emotion analysis in human-machine interaction," in *2015 10th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)*, 2015, pp. 1-6.
- [4] I. Kollia, N. Simou, A. Stafylopatis, and S. Kollias, "Semantic image analysis using a symbolic neural architecture," *Image Analysis & Stereology*, vol. 29, no. 3, pp. 159-172, 2010.
- [5] B. Glimm, Y. Kazakov, I. Kollia, and G. Stamou, "Lower and upper bounds for SPARQL queries over OWL ontologies," in *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.
- [6] D. Kollias, A. Tagaris, S. Stafylopatis, S. Kollias, and G. Tagaris, "Deep neural architectures for prediction in healthcare," *Complex & Intelligent Systems*, vol. 4, no. 2, pp. 119-131, 2018.
- [7] D. Kollias et al., "Adaptation of contextualization of deep neural network models," in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2017, pp. 1-8.
- [8] I. Kollia, A. Stafylopatis, and S. Kollias, "Predicting parkinson's disease using latent information extracted from deep neural networks," in *2019 IEEE International Joint Conference on Neural Networks (IJCNN)*, 2019, pp. 1-8.
- [9] J. Wingate, I. Kollia, L. Bidaut, and S. Kollias, "Unified deep learning approach for prediction of Parkinson's disease," *IET Image Processing*, 2020.
- [10] D. Kollias et al., "Deep transparent prediction through latent representation analysis," *European Conference on Artificial Intelligence, TAILOR Workshop*, pp. 1-16, August 2020.
- [11] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*.: MIT Press, 2016.
- [12] P. Sadjja, "Machine learning for detection and diagnosis of disease," *Annual Review of Biomedical Engineering*, pp. 537-565, 2006.
- [13] A. Tagaris, D. Kollias, A. Stafylopatis, G. Tagaris, and S. Kollias, "Machine learning for neurodegenerative disorder diagnosis - survey of practices and launch of benchmark dataset," *International Journal on Artificial Intelligent Tools*, vol. 27, no. 3, 2018.
- [14] Y. Zhang and A. Kagen, "Machine learning interface for medical image analysis," *Journal of Digital*

*Imaging*, vol. 30, no. 5, pp. 615-621, 2017.

- [15] A. Tagaris, D. Kollias, and A. Stafylopatis, "Assessment of parkinson's disease based on deep neural networks," in *Proceedings of the 17th International Conference on Engineering Applications of Neural Networks*, Athens, Greece, 2017, pp. 391-403.
- [16] [Online]. <https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9>
- [17] [Online]. <https://medium.com/greyatom/what-is-underfitting-and-overfitting-in-machine-learning-and-how-to-deal-with-it-6803a989c76>
- [18] [Online]. <https://www.geeksforgeeks.org/underfitting-and-overfitting-in-machine-learning/>
- [19] [Online]. <https://mlnotebook.github.io/post/CNN1/>
- [20] [Online]. <https://medium.com/abraia/getting-started-with-image-recognition-and-convolutional-neural-networks-in-5-minutes-28c1dfdd401>
- [21] [Online]. <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>
- [22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [23] [Online]. <https://towardsdatascience.com/understanding-and-coding-a-resnet-in-keras-446d7ff84d33>
- [24] C. Tan et al., "A survey on deep transfer learning," *27th International Conference on Artificial Neural Networks*, pp. 270-279, 2018.
- [25] D. Kollias and S. P. Zafeiriou, "Training deep neural networks with different datasets in-the-wild: The emotion recognition paradigm," in *International Joint Conference on Neural Networks (IJCNN)*, 2018, pp. 1-8.
- [26] [Online]. [https://pennylane.ai/qml/demos/tutorial\\_quantum\\_transfer\\_learning.html](https://pennylane.ai/qml/demos/tutorial_quantum_transfer_learning.html)
- [27] [Online]. <https://towardsdatascience.com/clustering-clearly-explained-5561642ec20c>
- [28] D. Arthur and S. Vassilvitskii, "K-means++: The advantages of careful seeding," in *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms*, 2007, pp. 1027-1035.
- [29] [Online]. <https://stackoverflow.com/questions/60312401/when-using-the-k-means-clustering-algorithm-is-it-possible-to-have-a-set-of-dat>
- [30] [Online]. <https://github.com/ieee8023/covid-chestxray-dataset>
- [31] [Online]. <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data>

- [32] [Online]. <https://github.com/UCSD-AI4H/COVID-CT>
- [33] [Online]. <https://www.medrxiv.org/content/10.1101/2020.04.13.20063941v1.full.pdf+html>
- [34] [Online]. [https://en.wikipedia.org/wiki/COVID-19\\_pandemic](https://en.wikipedia.org/wiki/COVID-19_pandemic)
- [35] [Online]. <https://www.cebm.net/covid-19/what-tests-could-potentially-be-used-for-the-screening-diagnosis-and-monitoring-of-covid-19-and-what-are-their-advantages-and-disadvantages/>
- [36] [Online]. <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>