



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

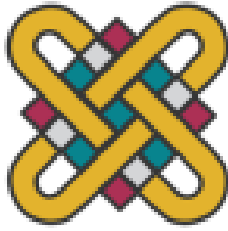
ΕΦΑΡΜΟΓΕΣ ΤΗΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΣΤΟΝ COVID-19



ΦΟΙΤΗΤΗΣ: ΣΤΑΥΡΟΣ ΣΑΜΑΡΑΣ (2162)

ΕΠΙΒΛΕΠΩΝ: ΚΑΘΗΓΗΤΗΣ ΣΙΝΑΤΚΑΣ ΙΩΑΝΝΗΣ

ΚΑΣΤΟΡΙΑ 2023



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΕΦΑΡΜΟΓΕΣ ΤΗΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΣΤΟΝ COVID-19

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

του

ΣΤΑΥΡΟΣ ΣΑΜΑΡΑΣ

(ΑΕΜ: 2162)

Επιβλέπων :

ΣΙΝΑΤΚΑΣ ΙΩΑΝΝΗΣ

ΚΑΘΗΓΗΤΗΣ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την

.....

Ον/μο Μέλους

Ιδιότητα Μέλους

.....

Ον/μο Μέλους

Ιδιότητα Μέλους

.....

Ον/μο Μέλους

Ιδιότητα Μέλους

Καστοριά Μάρτιος -2023

ΕΥΧΑΡΙΣΤΙΕΣ

Θέλω να ευχαριστήσω τον Καθηγητή, δάσκαλο και συνάνθρωπο μου Ιωάννη Σινάτκα, που δίχως την επιμέλειά του, τον επαγγελματισμό του, την επιμονή του και υπομονή του, προς το πρόσωπό μου, δε θα μπορούσα να συντάξω αυτήν την εργασία.

Copyright © 2022 – ΣΤΑΥΡΟΣ ΣΑΜΑΡΑΣ

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.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν αποκλειστικά τον συγγραφέα και δεν αντιπροσωπεύουν τις επίσημες θέσεις του Πανεπιστημίου Δυτικής Μακεδονίας.

Ως συγγραφέας της παρούσας εργασίας δηλώνω πως η παρούσα εργασία δεν αποτελεί προϊόν λογοκλοπής και δεν περιέχει υλικό από μη αναφερόμενες πηγές.

ΠΕΡΙΛΗΨΗ

Στη παρούσα εργασία έχουμε την παρουσίαση των εργασιών ομάδων για τη μελέτη με Τεχνητή Νοημοσύνη για τον Covid-19 στο πεδίο της διάγνωσης μέσω αξονικών τομογραφιών, στο πεδίο της φαρμακοσύνθεσης και της κατασκευής εμβολίου ενάντια στην πρωτεΐνη ακίδα του ιού.

Ειδικότερα, η ομάδα των Xu et al. με την ανέπτυξε μοντέλο ανίχνευσης αλλοιώσεων από τον Covid-19 σε CT εικόνες πνεύμονα. Το μοντέλο αυτό με το όνομα ResNet-18, το οποίο εφαρμόζει 3D τμηματοποίηση. Το μοντέλο εμφανίζει διακριτική ικανότητα, ικανότητα ταυτοποίησης/αναγνώρισης σε ποσοστό ακρίβειας 86,7%.

Παράλληλα στο ίδιο πεδίο η ομάδα Wang et al. ανέπτυξε μοντέλο τεχνητής νοημοσύνης για την μελέτη των CT εικόνων πνεύμονα για μελέτη της επικινδυνότητας της κατάστασης του ασθενούς και υποτροπής. Το μοντέλο αυτό έχει βάση EfficientNet-B4. Το μοντέλο κατάφερε στη πρόγνωση κινδύνου να αγγίξει το ποσοστό 95% επαλήθευσης με τα ιατρικά/κλινικά δεδομένα, μία πολύ μεγάλη εγκυρότητα. Καθώς επίσης 97% στον εντοπισμό των πάσχων περιοχών στις ct εικόνες.

Στο πεδίο των φαρμάκων η ομάδα Bung et al., ανέπτυξε μοντέλο τεχνητής νοημοσύνης που βασίζεται σε παραγωγικό μοντέλο, και στη συνέχεια εφαρμογή μεταφορά μάθησης και υποστηριζόμενης μάθησης, ώστε να επιλέξει το κατάλληλο μόριο αναστολέα πρωτεάσης του Covid-19, με βάση δομή και φυσικοχημικές ιδιότητες των μορίων και πρωτεϊνών.

Στο πεδίο της κατασκευής εμβολίου η ομάδα του Yang Z. et al. δημιούργησε το DeepVacPred, το οποίο υπολογίζει, τους επιτόπους στα T, B και Τκυτταροτοξικά λεμφοκύτταρα, ώστε να δημιουργηθεί εμβόλιο, ενάντια της ακίδας του Covid-19, πολλαπλών επιτόπων και άρα ισχυρή ανασοαπάντηση.

Λέξεις κλειδιά: Τεχνητή Νοημοσύνη, SAR-CoV-2, COVID-19

Abstract

The Covid-19 pandemic in 2020 to 2021, shocked the planet. Thousands of people dead, thousands of people were hospitalized in intensive care units. The global economy crash. In this situation, there was an immediate need for rapid diagnosis, new treatment, new vaccines and drugs. Artificial Intelligence was applied in these areas. This thesis presents these applications.

The team of Xu et al. developed a 3D learning model to analyse the CT images, recognise the infected area from Covid-19. This model use VNet-inception residual network (IR) technology with 3D segmentation process. The accuracy rate was 86.7% in terms of all CT case taken together. The team of Wang et al. developed An artificial intelligence (AI) system in a time-to-event analysis framework, to integrate chest CT and clinical data for risk prediction of future deterioration to critical illness in patient with Covid-19. This system has the architecture of EfficientB4. The AI system achieved a C-index of 0.80 for predicting individual COVID-19 patients' to critical illness. The AI system successfully stratified the patients into high-risk and low-risk groups with distinct progression risks.

The team of Bung et al. developed a deep neural network-based generative and predictive models for *de novo* design of small molecules capable of inhibiting the 3CL protease. The generative model was optimized using transfer learning and reinforcement learning to focus around the chemical space corresponding to the protease inhibitors. Multiple physicochemical property filters and virtual screening score were used for the final screening. The team of Yang et al. developed an in silico deep learning approach for prediction and design of a multi-epitope vaccine (DeepVacPred). By combining the in silico immunoinformatics and deep neural network strategies, the DeepVacPred computational framework directly predicts 26 potential vaccine subunits from the available SARS-CoV-2 spike protein sequence. This artificial intelligence (AI) based vaccine discovery framework accelerates the vaccine design process and constructs a 694aa multi-epitope vaccine containing 16 B-cell epitopes, 82 CTL epitopes and 89 HTL epitopes, which is promising to fight the SARS-CoV-2 viral infection and can be further evaluated in clinical studies.

Key words: Artificial intelligence, SARS-CoV-2, Covid-19

Περιεχόμενα

ΠΕΡΙΛΗΨΗ.....	7
Abstract.....	8
Πίνακας Περιεχομένων Εικόνων.....	12
Εισαγωγή.....	14
Πρώτο μέρος Covid-19.....	14
Η βιολογία του ιού covid-19.....	14
Ο μηχανισμός μόλυνσης.....	17
Η επιδημία.....	18
Παθολογοανατομία Ευρήματα.....	21
Δεύτερο μέρος Τεχνητή Νοημοσύνη.....	22
Τεχνητή Νοημοσύνη - Ορισμός Ιστορική αναδρομή.....	22
Μηχανική Μάθηση – Τεχνητά Νευρωνικά Δίκτυα.....	23
Αλγόριθμοι Τεχνητής Νοημοσύνης – Μηχανικής Μάθησης.....	24
Αλγόριθμος Random Forest.....	24
Ενισχυτική Μάθηση.....	25
Τύποι Νευρωνικών Δικτύων – Μηχανικής Μάθησης.....	26
Αναδρομικό νευρωνικό δίκτυο.....	26
Βαθύ νευρωνικό δίκτυο.....	27
Συνελεκτικό Νευρωνικό Δίκτυο.....	28
Μηχανές διανυσμάτων υποστήριξης.....	29
Deep Convolutional Neural Fields.....	30
ResNet τοπολογία.....	31
V-Net.....	32
3D U-net.....	34
VNet-inception-residual network.....	35
EfficientNet B4.....	36
Στατιστικές μέθοδοι επαλήθευσης/ανάλυσης.....	37
Στατιστική μέθοδο Noisy-OR Bayesian.....	37
Διασταυρούμενη εντροπία.....	38
Wald test.....	39
C-Index.....	39
AMSGrad.....	40

Adam	40
Κύριο Μέρος.....	41
Η Τεχνητή Νοημοσύνη στον covid-19	41
Εφαρμογές σε CT-X ακτινογραφίες	41
Μοντέλο ομάδας Xu et al.....	41
Δεδομένα/Υλικά	41
Η διαδικασία επεξεργασίας των CT image	41
Δημιουργία μοντέλου μάθησης	42
Εντοπισμός περιοχής	43
Δομή Νευρωνικού Δικτύου	43
Υλικό εφαρμογής (Hardware)	45
Διαδικασία εκπαίδευσης.....	45
Αποτελέσματα	46
Μοντέλο ομάδας Wang et al.	47
Δεδομένα Υλικά	47
Μοντέλο Τεχνητής Νοημοσύνης.....	47
Μεθοδολογία διαχωρισμού πρώτου σταδίου	49
Μεθοδολογία μοντέλου πρόγνωσης κινδύνου	49
Μέθοδος Πρόβλεψης Πρόγνωσης.....	50
Στατιστική ανάλυση	51
Υλικό εφαρμογής (Hardware)	51
Απεικόνιση απόδοσης μοντέλου	52
Αποτελέσματα	52
Εφαρμογές στην ανακάλυψη φαρμάκων	53
De novo σχεδιασμός χημικών μορίων ενάντια στον SARS-CoV-2	53
Συλλογή δεδομένων.....	53
Προεπεξεργασία δεδομένων.....	53
Μαθαίνοντας την γλώσσα των μικρών μορίων με τη χρήση παραγόμενων μοντέλων....	54

Σχεδιασμός φαρμάκων ligand-based χρησιμοποιώντας μεταφορά μάθησης.....	55
Βελτιστοποίηση ιδιοτήτων με βάση το μοντέλο πρόβλεψης	55
Συνδυασμός του παραγόμενου και προγνωστικού μοντέλου με τη χρήση ενισχυτικής μάθησης	56
Φιλτράρισμα των παραγόμενων μορίων μέσω φυσικοχημικών ιδιοτήτων.....	57
Εφαρμογή φίλτρων για αφαίρεση μορίων με ανεπιθύμητες ομάδες.....	57
Μοριακή μοντελοποίηση.....	57
Αποτελέσματα	58
Εφαρμογή στην ανακάλυψη εμβολίων	60
DeepVacPred.....	60
Δεδομένα / υλικά	62
Οργάνωση Δεδομένων.....	62
Εκπαίδευση Δικτύου	63
Επαλήθευση.....	63
Ανάλυση Μοντέλου.....	65
Γραμμική πρόβλεψη επιτόπων Β-κυττάρων	66
Πρόβλεψη επιτόπων Τ κυτταροτοξικών λεμφοκυττάρων.....	66
Πρόβλεψη επιτόπων Τ βοηθητικών λεμφοκυττάρων.....	66
Πρόβλεψη δευτεροταγούς δομής	67
Εμβολίου τρισδιάστατο (3D) μοντέλο δομής.....	67
Επιβεβαίωση της 3D δομής του εμβολίου.....	67
Αποτελέσματα	67
Συμπεράσματα Συζήτηση	68
Αναφορές	73

Πίνακας Περιεχομένων Εικόνων

Εικόνα 1 Η πρωτεΐνη ακίδα και το γονιδίωμα του ιού	15
Εικόνα 2 Ιική μόλυνση κυττάρου και κύκλος ιικού πολλαπλασιασμού.....	16
Εικόνα 3 Σύνολο κρουσμάτων/ασθενών παγκοσμίως.....	18
Εικόνα 4 Σύνολο θανάτων παγκοσμίως.....	18
Εικόνα 5 Πορεία κρουσμάτων Ελλάδας.....	19
Εικόνα 6 Ημερήσια πορεία κρουσμάτων Ελλάδας.....	20
Εικόνα 7 Πορεία θανάτων Ελλάδας.....	20
Εικόνα 8 Ημερήσιοι θάνατοι Ελλάδας.....	20
Εικόνα 9 Το πλαίσιο της ενισχυτικής μάθησης.....	25
Εικόνα 10 Τοπολογία RNN.....	26
Εικόνα 11 Διαφορά μηχανικής μάθησης και Βαθιάς Μάθησης.....	27
Εικόνα 12 Δομή CNN.....	28
Εικόνα 13 Εφαρμογή ΜΔΥ.....	29
Εικόνα 14 Τυπικό CNF (A) και DeepCNF (B).....	30
Εικόνα 15 Αρχιτεκτονική του DeepCNF.....	30
Εικόνα 16 Αρχιτεκτονική του ResNet-50.....	31
Εικόνα 17 Αρχιτεκτονική Vnet.....	32
Εικόνα 18 Μεθοδολογία Downsampling (αριστερά) και Upsampling (δεξιά).....	33
Εικόνα 19 Αρχιτεκτονική 3D U-Net.....	34
Εικόνα 20 Αρχιτεκτονική του VNET-IR-RPN.....	35
Εικόνα 21 EfficientNet.κλιμάκωση.....	36
Εικόνα 22 Πύλη Noisy-OR Bayesian.....	37
Εικόνα 23 Διασταυρούμενη Εντροπία και Συνάρτηση Softmax.....	38
Εικόνα 24 Αποτελέσματα σε πιθανότητες και τα αναμενόμενα αποτελέσματα.....	38
Εικόνα 25 Μαθητικός τύπος Wald test.....	39
Εικόνα 26 Η στατιστική περιοχή είναι ίση με περιοχή κάτω από τη καμπύλη ROC.....	39
Εικόνα 27 Σύγκριση μεθόδων.....	40
Εικόνα 28 Μαθηματικός τύπος Wald test.....	40
Εικόνα 29 Σύγκριση μεθόδων.....	40
Εικόνα 30 Μαθηματικός τύπος Adam.....	40
Εικόνα 31 Στάδια επεξεργασίας εικόνων.....	42

Εικόνα 32 Περιοχές εστιών μόλυνσης (α) Covid-19 ground-glass , (b) IAVP εστίες, (c) ελάχιστη απόσταση μάσκας κέντρου φλεγμονής, (d) Διαγώνιος ελάχιστου περιγεγραμμένου ορθογωνίου	43
Εικόνα 33 Σχηματική αναπαράσταση νευρωνικού δικτύου	44
Εικόνα 34 Η πορεία μείωσης της Loss σε Εποχές.....	45
Εικόνα 35 Η πορεία βελτίωσης ακρίβειας σε Εποχές.....	45
Εικόνα 36 Παρουσίαση επιλογής ασθενών.	47
Εικόνα 37 Παρουσίαση διαχωρισμού CT εικόνων. Manual (red) and automatic (blue) lung (top row) and lesion (bottom row) segmentations on 3 slices.....	48
Εικόνα 38 Περιγραφή του μοντέλου Τεχνητής Νοημοσύνης στην επεξεργασία και εκπαίδευση του.....	48
Εικόνα 39 Απεικόνιση απόδοσης μοντέλου.	52
Εικόνα 40 De novo βήματα σχεδιασμού φαρμάκων, για τη δημιουργία μικρών μορίων ενάντια σε στόχο.....	54
Εικόνα 41 Βαθμολόγηση μορίων και ομαδοποίηση σε σχέση με καταστολείς πρωτεάσης του HIV.	58
Εικόνα 42 Τελική επιλογή και εφαρμογή των μορίων/καταστολέων, σε σχέση με την δράση τους στη πρωτεάση του Covid-19.	59
Εικόνα 43 Σχεδιάγραμμα της Υπολογιστικής Σχεδίασης εμβολίων/επιτόπων.....	61
Εικόνα 44 Καμπύλη επαλήθευσης μοντέλου.....	64
Εικόνα 45 Σύγκριση μοντέλων TN.....	71

Εισαγωγή

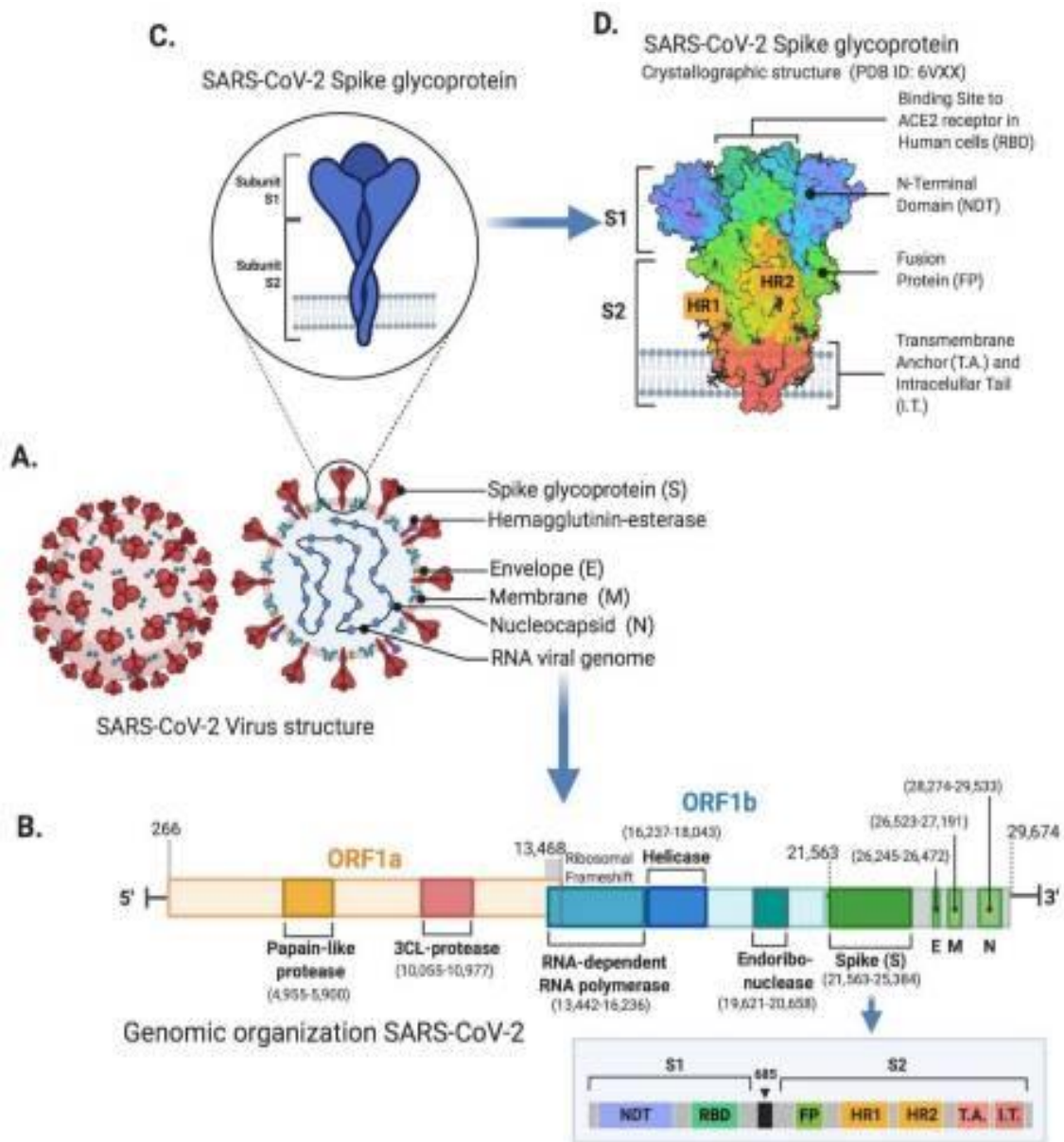
Πρώτο μέρος Covid-19

Η βιολογία του ιού covid-19

Ο ιός SARS-CoV-2 ανήκει στο γένος των ιών Betacoronavirus και είναι μέλος της οικογένειας των Coronavirinae (Rehman S.U., 2020). Ο ιός έχει σχήμα σχεδόν σφαιρικό, με διάμετρο 60-140 nm. Οι κορωνοϊοί έχουν ένα μεγάλο μονόκλωνο RNA ως γενετικό υλικό, με μήκος 27-32 χιλιάδες βάσεις (Kb). (Εικόνα 1) (Hobman T.C., 2020) (Chilamakuri R., 2021).

Στην οικογένεια αυτή, έχει διαπιστωθεί ότι, αρκετοί κορωνοϊοί κωδικοποιούν στο γονιδίωμα τους κωδικοποιούν πρωτεΐνες όπως η εστεράση της αιματογλουτίνης (hemagglutinin-esterase), την 3a/b πρωτεΐνη και την 4a/b πρωτεΐνη, τις οποίες και εκφράζουν/εμφανίζουν στην επιφάνειά τους (Chilamakuri R., 2021). Ειδικότερα στον SARS-CoV-2, το γονιδίωμα του εμφανίζει ομοιότητα με τους υπόλοιπους κορωνοϊούς, το οποίο αποτελείται κυρίως από ανοικτά πλαίσια ανάγνωσης (open reading frames – ORFs). Περίπου το 67% κωδικοποιείται από το ORFa/b και κωδικοποιεί 16 μη δομικές πρωτεΐνες (nonstructural -nsp1-16), ενώ το υπόλοιπο 33% κωδικοποιεί πρωτεΐνες για τη μόλυνση καθώς και δομικές πρωτεΐνες. Οι ORF1a και ORF1b περιέχουν μετατοπισμένα πλαίσια (frameshift) τα οποία παράγουν 2 πολυπεπτίδια τα pp1a και pp1ab. Παράλληλα, κωδικοποιούνται από το 16nps, τη πρωτεάση της παπαΐνης (PLpro) ή τη πρωτεάση της χυμοθρυψίνης (3CLpro) (Εικόνα 1B) (Liu Y., 2020).

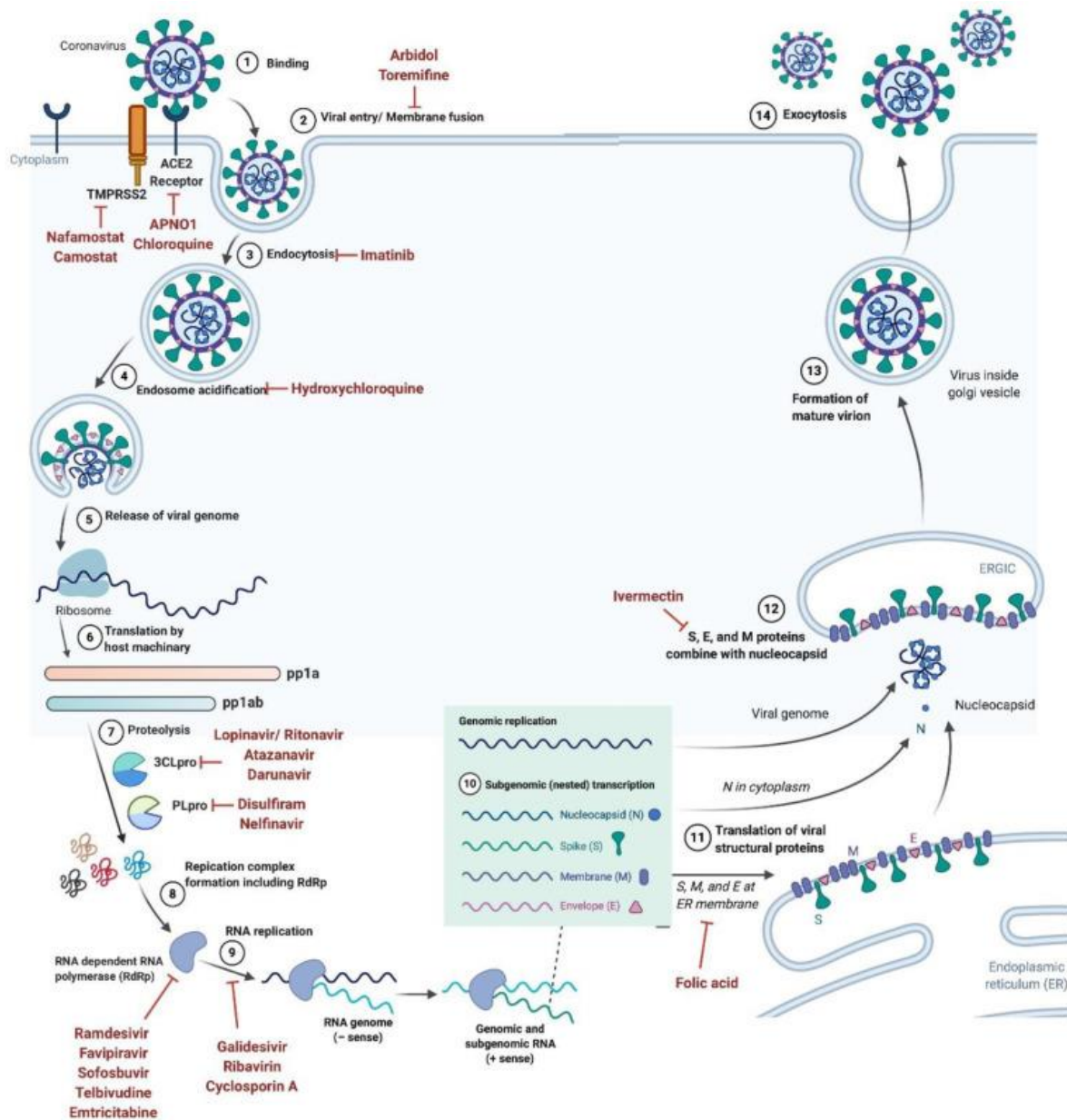
Ο SARS-CoV-2, όσο αναφορά δομικές πρωτεΐνες, κωδικοποιεί τουλάχιστον 4 κύριες πρωτεΐνες και ειδικότερα την ακίδα (spike - S), τη μεμβρανική πρωτεΐνη M, την πρωτεΐνη του ιικού φακέλου E, και την πρωτεΐνη του νουκλεοψαγιδίου N. Αυτές αναφέρονται συχνά με τα προηγούμενα αρχικά γράμματα, τα οποία χρησιμοποιούνται για να αναφερθούμε και στα αντίστοιχα γονιδιά τους, στο ORF 10 και 11, τα οποία αποτελούν το 1/3 του γονιδιώματος κοντά στο 3'-άκρο (Εικόνα 2 A,B) (Mousavizadeh L., 2020). Αυτές οι ώριμες πρωτεΐνες είναι υπεύθυνες για τον ιική αντιγραφή (Cui J., 2019). Οι περισσότεροι ανιχνευτές (probes) και εκκινητές για την ανίχνευση του SARS-CoV-2 είναι κατασκευασμένοι για τον εντοπισμό του ORF1ab και του N γονιδίου (Mollaei H.R., 2020).



Εικόνα 1 Η πρωτεΐνη ακίδα και το γονιδίωμα του ιού

Όταν ο ιός εισέρχεται σε ένα κύτταρο ξενιστή, θα ακολουθήσει η σύνθεση των πρωτεϊνών του. Η σύνθεση και πολλαπλασιασμός του RNA, γίνεται με τη βοήθεια της RNA πολυμεράσης η οποία και αντιγράφει αξιόπιστα το αρχικό μόριο (Εικόνα 2) (Hobman T.C., 2020) (Sanders W., 2020). Για την επιτυχημένη πρόσδεση στο κύτταρο, ο SARS-CoV-2, «βλέπει» τον υποδοχέα ACE2 (Angiotensin-converting enzyme 2), ο οποίος υποδοχέας εμφανίζει πρόσδεση ισχυρότερη σε σχέση με άλλους SARSs ιούς, και εμφανίζει υψηλή μεταλλαξιγένεση στον ιό αυτό (Gussow A.B., 2020) (Chilamakuri R., 2021). Η M πρωτεΐνη είναι στο καψίδιο, ως γλυκοπρωτεΐνη, και εμφανίζεται σε μεγαλύτερη ποσότητα. Η M, είναι υπεύθυνη για τη μεταφορά θρεπτικών ουσιών από τη κυτταρική μεμβράνη καθώς και δίνει σχήμα στον ιό (Schoeman D., 2019). Η πρωτεΐνη ακίδα (S), είναι τύπου I γλυκοπρωτεΐνη και αποτελείται από

έτερα μέρη. Η N πρωτεΐνη βοηθάει στη σύνδεση του ιικού RNA καθώς και αποτελεί παράγοντας που συμβάλει στη σταθερότητά του (Kim D., 2020). Η E πρωτεΐνη διαδραματίζει σημαντικό ρόλο στην απελευθέρωση και ως αποτέλεσμα στη παθογένεση (Εικόνα 1 και 2) (Chilamakuri R., 2021). Η ανάλυση του γενετικού υλικού του SARS-CoV-2, έδειξε ότι σε ποσοστό 85%-95% της αλληλουχίας του αποτελεί όμοιο με τον SARS-CoV (Chen Y. L. Q., 2020).



Εικόνα 2 Ιική μόλυνση κυττάρου και κύκλος ιικού πολλαπλασιασμού.

Ο μηχανισμός μόλυνσης

Στη διαδικασία της μόλυνσης του ιού, στο κύτταρο ξενιστή, δηλαδή της εισαγωγής του σε αυτό, κομβικό ρόλο διαδραματίζει η πρωτεΐνη ακίδα (S) της επιφάνειας του ιού. Η πρωτεΐνη ακίδα είναι αυτή που δίνει και το σχήμα κορώνας στην οικογένεια αυτή των ιών (Εικόνα 1 C,D) (Chilamakuri R., 2021).

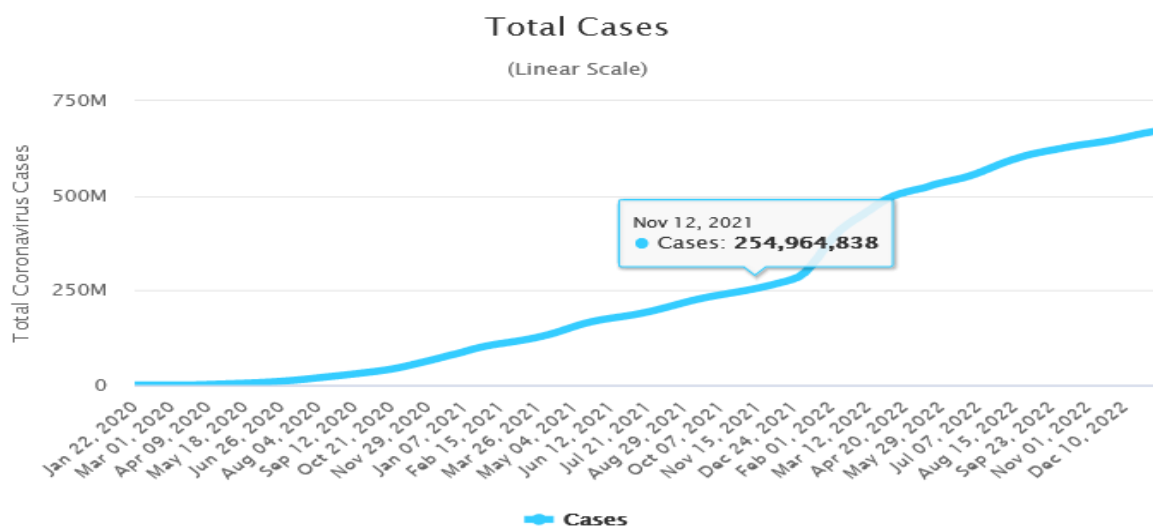
Η πρωτεΐνη ακίδα αποτελείται από 3 μέρη, ένα οκταμερές, δηλαδή σύμπλοκο 8 έτερων πρωτεϊνών, μία «άγκυρα» διαμεμβρανικής μεταφοράς και μια ενδοκυτταρική C-άκρο ουρά (Li F., 2016). Το οκταμερές, δομείται από τη περιοχή S1, που συνδέεται με τον υποδοχέα και την S2 περιοχή, η οποία δρα για τη σύντηξη με τη κυτταρική μεμβράνη (Εικόνα 1) (Mercurio I., 2020).

Η είσοδος στο κύτταρο ξενιστή, γίνεται με την αλληλεπίδραση/πρόσδεση της S1 περιοχής με τον υποδοχέα ACE2 (Mahmoud I.S., 2020). Ο ανθρώπινος υποδοχέας ACE2, εκφράζεται σχεδόν σε όλους τους ιστούς και κυρίως στα κύτταρα των πνευμόνων, των νεφρών, του εγκεφαλικού στελέχους, του λιπώδους ιστού, της καρδιάς, των αγγείων, του ήπατος, του στομάχου, καθώς και του επιθηλίου της μύτης και του στόματος (Xu H., 2020).

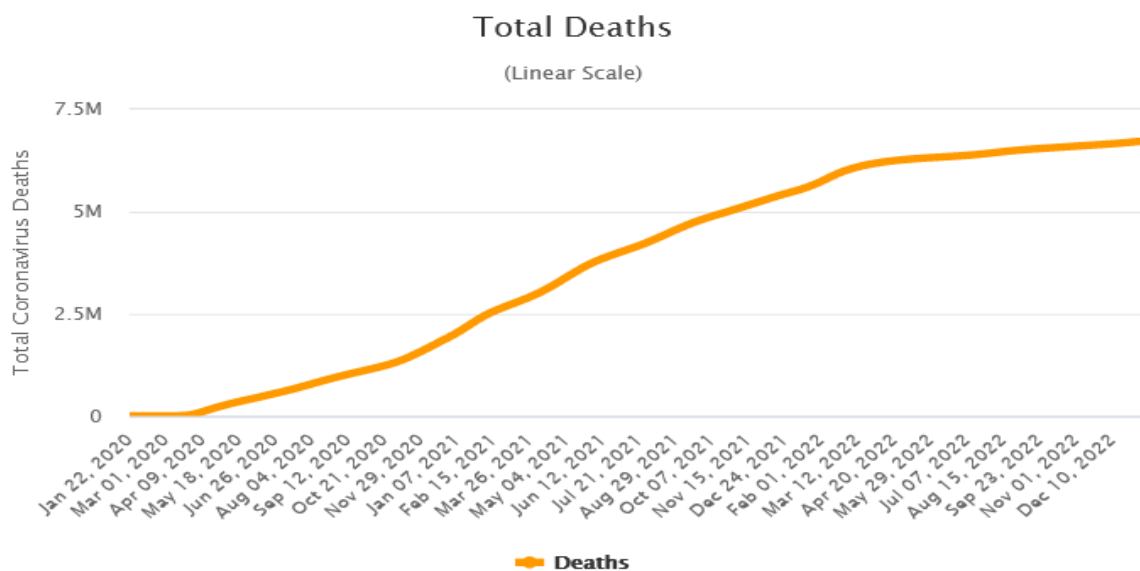
Η S2 υπομονάδα συγχωνεύει τις μεμβράνες του ιού και του κυττάρου-ξενιστή, καθώς επίσης βοηθάει στην είσοδο του ιικού γενετικού υλικού στο κύτταρο (Chen Y. G. Y., 2020). Αυτή η διαδικασία απαιτεί διάσπαση των S1 και S2 περιοχών, κάτι που γίνεται με δράση κασπασών από τον ξενιστή. Ειδικότερα διαπιστώθηκε ότι το ο SARS-CoV-2 χρησιμοποιεί τον ACE2 για την είσοδο, και στη συνέχεια την διαμεμβρανική πρωτεάση σερίνη 2 (Transmembrane protease serine 2 – TMPRSS2) και την ενδοσωμική πρωτεάση κυστεΐνης καθεψίνη B και L (endosomal cysteine protease serine B and L – CatB/L) (Εικόνα 1). Τα βήματα αυτά είναι σημαντικά ως σημεία δράση φαρμάκων αναστολής του ιού (Hoffmann M., 2020).

Η επιδημία

Σύμφωνα με την εβδομαδιαία επιδημιολογική ανασκόπηση/καταγραφή του Παγκοσμίου Οργανισμού Υγείας (ΠΟΥ) μέχρι τη 12^η Ιανουαρίου 2023 υπήρχαν 670.327.323 κρούσματα (ασθενείς) (Εικόνα 3) και 6.725.056 θάνατοι που είχα επιβεβαιωθεί (Εικόνα 4). Παράλληλα, παγκοσμίως είχαν γίνει 13.214.793.043 δόσεις εμβολίων.



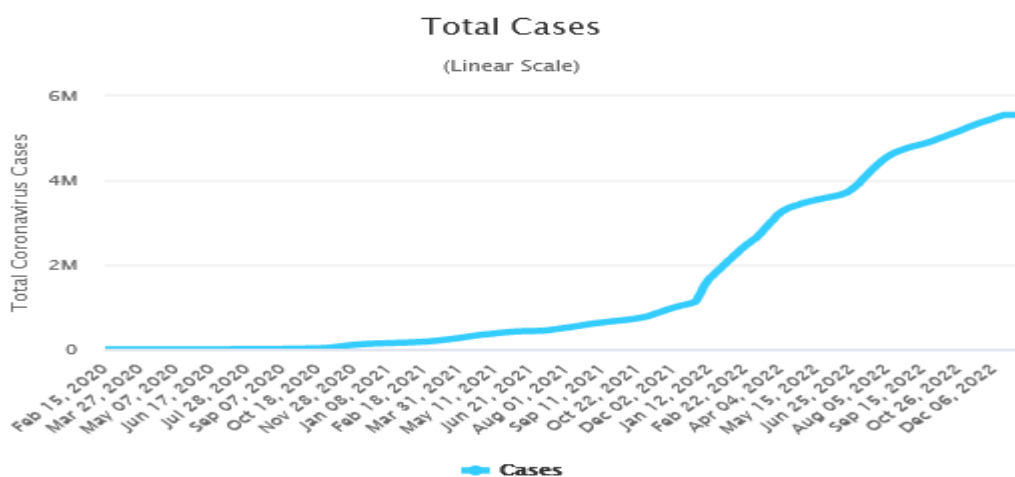
Εικόνα 3 Σύνολο κρουσμάτων/ασθενών παγκοσμίως.



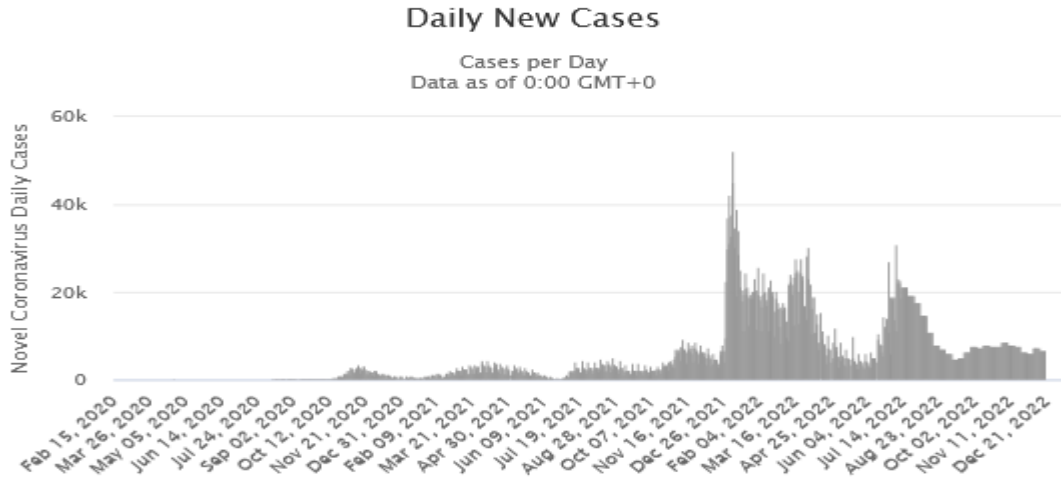
Εικόνα 4 Σύνολο θανάτων παγκοσμίως

Στην Ελλάδα, έως την ίδια ημερομηνία υπήρξαν 5.548.487 κρούσματα (Εικόνα 5) και 34.779 θάνατοι (Εικόνα 7). Παράλληλα είχα πραγματοποιηθεί 22.053.622 δόσεις εμβολίων. Σχετικά με την πορεία της πανδημίας στην Ελλάδα μέσα από τα γραφήματα (Εικόνα 5,6) βλέπουμε ότι η κορύφωση των ημερήσιων κρουσμάτων έγινε το διάστημα Δεκεμβρίου 2021 – Ιανουαρίου 2022, όπου το μέγιστο ημερήσιο ρεκόρ κρουσμάτων ανήλθε σε 52.098 κρούσματα (ειδικότερα την 4^η Ιανουαρίου 2022). Στο πλαίσιο των θανάτων (Εικόνα 7,8), βλέπουμε 3 κύρια κύματα, όπου το πρώτο ξεκινά από τα τέλη Οκτωβρίου 2020 – Δεκέμβριος 2020, το δεύτερο κύμα Μάρτιος 2021 – Μάιος 2021 και τέλος μία αρκετά εκτενής περίοδος υψηλού αριθμού καθημερινών θανάτων από Νοέμβριο 2021 – Απρίλιος 2022.

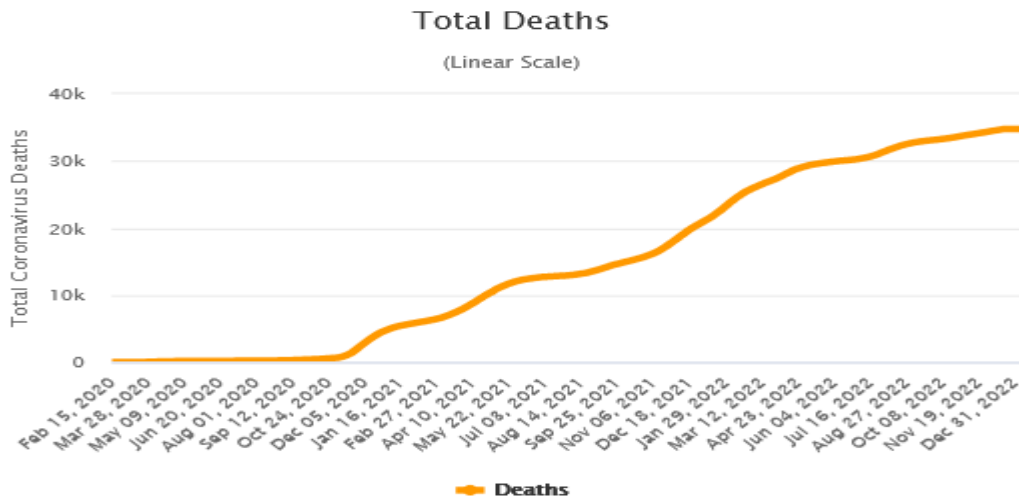
Όσο αναφορά την εικόνα των ασθενών, σε διεθνής μελέτες, πρόκειται για μία ακόμα μη σταθμισμένη πλήρως της παθολογικής εικόνας, ωστόσο έχει παρατηρηθεί το 14% ανέπτυξε σοβαρή συμπτωματολογία, όπου έπρεπε να νοσηλευτεί με την παροχή οξυγόνου και το 5% των ασθενών εισαγωγή σε μονάδα εντατικής θεραπείας (Vital., 2020). Στα πλαίσια του Covid-19 η κλινική εικόνα, στα σοβαρά περιστατικά, περιλάμβανε την ανάπτυξη Σύνδρομο Οξείας Αναπνευστικής Δυσχέρειας (Acute Respiratory Distress Syndrome – ARDS), σηπτικό σοκ καθώς και σήψη, καθώς και πολυοργανική ανεπάρκεια όπως βλάβες στον καρδιακό ιστό και οξεία νεφρική ανεπάρκεια (Yang X., 2020). Αυτή η τόσο σοβαρή και πολυεστιακή και πολυεπίπεδη ιατρική εικόνα, εμφανίζεται να συνδέεται και με την ηλικία, όπου η ηλικιωμένοι είναι σε δεινότερη θέση, καθώς και οι ασθενής με βεβαρημένο ιατρικό ιστορικό όπως καρδιακή ανεπάρκεια, υψηλή αρτηριακή πίεση, διαβήτη ή/και αναπνευστικά προβλήματα (Yang J., 2020).



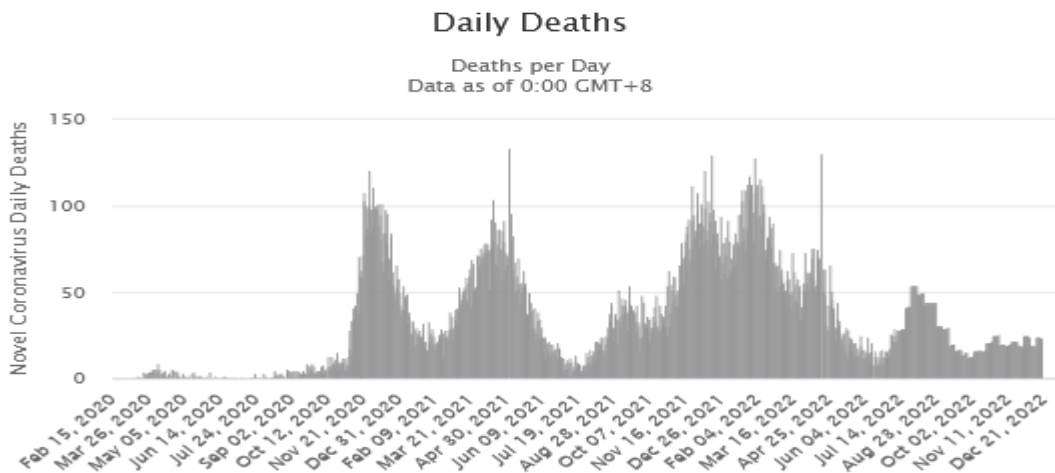
Εικόνα 5 Πορεία κρουσμάτων Ελλάδας.



Εικόνα 6 Ημερήσια πορεία κρουσμάτων Ελλάδας.



Εικόνα 7 Πορεία θανάτων Ελλάδας.



Εικόνα 8 Ημερήσιοι θάνατοι Ελλάδας.

Παθολογοανατομία Ευρήματα

Η εκδήλωση των συμπτωμάτων από τη μόλυνση (φάση επώασης) του Covid-19 εμφανίζεται περίπου στη 5^η μέρα (μέσος όρος χρόνου 5.1 ημέρες) (Shereen MA, 2020) και όσοι φορείς εκδηλώσουν συμπτώματα, η χρονική περίοδο τους είναι 11,5 ημέρες (Lauer SA, 2020). Τα χρονικά αυτά διαστήματα, ωστόσο, είναι στενά συνδεδεμένα με τη κατάσταση του ανοσοποιητικού συστήματος του φορέα/ασθενούς όπως και της ηλικίας του (Mohamadian M., 2021).

Στο πεδίο των γαστρεντερικών διαταραχών εντάσσονται η διάρροια, ο έμετος, η έλλειψη όρεξης. Τα συμπτώματα αυτά εμφανίζονται στο 40% του συνόλου των ασθενών (Adhikari SP, 2020) (J-J Z, 2020), ενώ από αυτούς με τα γαστρεντερικά συμπτώματα, ένα 10% δεν παρουσιάζει εικόνα εμπύρετης κατάστασης ή αναπνευστικής διαταραχής (λοίμωξης) (Wang D, 2020). Παράλληλα, ο Covid-19 έχει συνδεθεί με τη διαταραχή της πήξης του αίματος και ειδικότερα ανάπτυξης υπερπηκτικότητας του αίματος, η οποία οδηγεί σε αυξημένο κίνδυνο ανάπτυξης θρόμβων στη κυκλοφορία του αίματος (Danzi GB, 2020) (Mohamadian M., 2021).

Στο πλαίσιο των νευρολογικών αναλύσεων, βρέθηκε ότι ο ιός προσβάλλει και το νευρικό ιστό, με αποτέλεσμα ανάπτυξη συμπτωμάτων όπως κόπωση, ζάλη, διαταραχή επίγνωσης αλλά και σοβαρότερα όπως ισχαιμικά και αιμορραγικά εγκεφαλικά επεισόδια όπως και αλλοιώσεις σε μυϊκό ιστό (μυϊκά προβλήματα) (Mao L, 2020). Στο πλαίσιο του δερματικού ιστού, ενδιαφέρουσα είναι η παρατήρηση δερματικών αλλοιώσεων στο 20% των ασθενών στην Ιταλία (S., 2020).

Στα παιδιά, η κλινική πρόγνωση, σχετίζεται με την ανάπτυξη ή όχι σοβαρής αναπνευστικής ανεπάρκειας και κατά πόσο αυτή μπορεί να αντιμετωπιστεί με χορήγηση οξυγόνου, εντός διαστήματος 1-3 ημερών, στις σοβαρές περιπτώσεις (Salehi S, 2020) (Mohamadian M., 2021). Χαρακτηριστικά γνωρίσματα είναι το σηπτικό σοκ, η σήψη, η ακατάπαυστη και διαρκής αιμορραγία και η μεταβολική οξέωση (Chen Z-M, 2020). Το σηπτικό σοκ μπορεί να προκαλέσει πολυοργανική βλάβη πέρα από την ήδη αναπνευστική ανεπάρκεια. Όταν εμφανίζεται η πολυοργανική διαταραχή, σε συνδυασμό με τις διαταραχές πήξης και το πιθανό σηπτικό σοκ, ο κίνδυνος θανάτου αυξάνεται δραματικά (Chen Z-M, 2020). Τέλος, τα παιδιά είναι επιρρεπή σε μία υπερφλεγμονώδη απόκριση του ανοσοποιητικού έναντι του Covid-19 παρόμοια με τη νόσο Kawasaki η οποία μπορεί να αντιμετωπιστεί (Loomba RS, 2020).

Επίσης ενδιαφέρουσα είναι και η ηλικιακή κατανομή όπου η διάμεση ηλικία των ασθενών είναι τα 59 έτη σε ένα εύρος ηλικιών από 15 ως 89 (Adhikari SP, 2020).

Δεύτερο μέρος Τεχνητή Νοημοσύνη

Τεχνητή Νοημοσύνη - Ορισμός Ιστορική αναδρομή

Η ικανότητα του ανθρώπου, η δυνατότητα του ανθρώπινου εγκεφάλου, να διαθέτει σκέψη, συνείδηση, αυτογνωσία αποτελεί απόρροια, ενός βαθέως και πολύπλοκου νευρωνικού δικτύου, το οποίο αναπτύχθηκε τόσο με την επίδραση τις βιολογικής εξέλιξης των ειδών, όσο και με την ικανότητά του, να προσπαθεί να κατανοήσει/να μάθει και να ανταποκριθεί στο περιβάλλον.

Με την ανάπτυξη των υπολογιστών, ο άνθρωπος σκέφτηκε την ανάπτυξη υπολογιστικών δομών (υλικού & λογισμικού), ώστε να εμφανιστεί η νοημοσύνη που αυτός διαθέτει, δηλαδή η τεχνητή νοημοσύνη. Ορισμός για την τεχνητή νοημοσύνη μπορεί να αποτελέσει:

«Τεχνητή Νοημοσύνη είναι ο τομέας της Επιστήμης των Υπολογιστών που ασχολείται με τη σχεδίαση και την υλοποίηση προγραμμάτων τα οποία είναι ικανά να μιμηθούν τις ανθρώπινες γνωστικές ικανότητες, εμφανίζοντας έτσι χαρακτηριστικά που αποδίδουμε συνήθως σε ανθρώπινη συμπεριφορά, όπως για παράδειγμα η επίλυση προβλημάτων, η αντίληψη μέσω της όρασης, η μάθηση, η εξαγωγή συμπερασμάτων η κατανόηση φυσικής γλώσσας, κτλ.» (Βλαχάβας Ι., 2011 (Γ' Έκδοση))

Ως προσεγγίσεις, που έχουν υπάρξει, για τη τεχνητή νοημοσύνη, αυτές είναι δυο κλάδων:

- Η κλασσική ή συμβολική τεχνητή νοημοσύνη, στην οποία έχουμε τη κατανόηση των διεργασιών της νοημοσύνης και ασχολείται με τη προσομοίωση αυτής, μέσω της ανάπτυξης αλγορίθμων και συστημάτων που βασίζονται στη γνώση, χρησιμοποιώντας ως δομικές μονάδες τα σύμβολα.
- Η υπολογιστική νοημοσύνη ή συνδετική ή μη συμβολική τεχνητή νοημοσύνη, η οποία βασίζεται στο να μιμηθεί τις βιολογικές διεργασίες, όπως η εξέλιξη των βιολογικών ειδών, η λειτουργία του εγκεφάλου.

Ακρογωνιαίος λίθος, για την τεχνητή νοημοσύνη, αποτελεί το Turing test, το οποίο αναπτύχθηκε το 1953, από τον Alan Turing, πατέρα της Τεχνητής Νοημοσύνης και σύμφωνα με το οποίο, εάν εφαρμοστεί σειρά ερωτημάτων σε άνθρωπο και σύστημα Τεχνητής Νοημοσύνης και ο εξεταστής δε μπορεί εκ των απαντήσεων να προσδιορίσει ποιος είναι τί, τότε έχουμε σύστημα με επίπεδο τεχνητής νοημοσύνης (Βλαχάβας Ι., 2011 (Γ' Έκδοση)).

Μηχανική Μάθηση – Τεχνητά Νευρωνικά Δίκτυα

Ο ανθρώπινος εγκέφαλος κατανοεί το περιβάλλον του μέσω της παρατήρησης και της δημιουργίας μίας αφαιρετικής εκδοχής αυτού, δηλαδή ενός μοντέλου. Η δημιουργία ενός τέτοιου μοντέλου, καλείται επαγωγική μάθηση, και η διαδικασία επαγωγή. Παράλληλα, ο ανθρώπινος εγκέφαλος, έχει την ικανότητα να οργανώνει, να εξάγει συσχετίσεις των διαφόρων εμπειριών και των καταστάσεων του, και να δημιουργεί νέες δομές που καλούνται πρότυπα. Η δημιουργία μοντέλων ή προτύπων από ένα σύνολο δεδομένων, σε ένα υπολογιστικό σύστημα, καλείται μηχανική μάθηση (Βλαχάβας Ι., 2011 (Γ' Έκδοση)).

Έχουν προταθεί διάφοροι ορισμοί για τη μηχανική μάθηση, χαρακτηριστικός είναι του Mitchell (1997), «ένα πρόγραμμα υπολογιστή θεωρείται ότι μαθαίνει από την εμπειρία E σε σχέση με μια κατηγορία εργασιών T και μια μετρική απόδοσης P , αν η απόδοση σε εργασίες της T , όπως μετριούνται με την P , βελτιώνονται με την εμπειρία E .».

Τα είδη της μηχανικής μάθησης διακρίνονται σε:

- Μάθηση με επίβλεψη ή μάθηση με παραδείγματα. Σε αυτήν την κατηγορία αυτή ανήκει η μάθηση εννοιών, τα δένδρα ταξινόμησης ή απόφασης, η μάθηση κανόνων, η μάθηση κατά περίπτωση, η μάθηση κατά Bayes, η γραμμική παρεμβολή, τα νευρωνικά δίκτυα και οι μηχανές διανυσμάτων υποστήριξης.
- Μάθηση χωρίς επίβλεψη ή μάθηση με παρατήρηση. Στην κατηγορία αυτή ανήκουν οι κανόνες συσχέτισης και οι ομάδες.

Ειδικότερα, τα τεχνητά νευρωνικά δίκτυα (ΤΝΔ) είναι συστήματα επεξεργασίας δεδομένων, αποτελούνται από πλήθος τεχνητών νευρώνων, οργανωμένων σε δομές. Παράλληλα υπάρχει οργάνωση σε επίπεδα/στρώματα. Το πρώτο ορίζεται επίπεδο εισόδου, το τελευταίο είναι το επίπεδο εξόδου, ενώ τα ενδιάμεσα επίπεδα ορίζονται ως κρυφά επίπεδα. Οι νευρώνες στα ΤΝΔ εμφανίζονται πλήρως συνδεδεμένοι, δηλαδή όλοι συνδέονται με όλους και οι μερικούς συνδεδεμένοι, όπου οι νευρώνες ενός επιπέδου είναι πλήρως συνδεδεμένοι με τους νευρώνες του επόμενου επιπέδου. Όταν δεν υπάρχουν συνδέσεις μεταξύ νευρώνων ενός επιπέδου και του προηγούμενου επιπέδου, έχουμε δίκτυα πρόσθιας τροφοδότησης. Στην αντίθετη περίπτωση έχουμε σύνδεση νευρώνων του ίδιου επιπέδου, και χαρακτηρίζονται ανατροφοδότηση (Βλαχάβας Ι., 2011 (Γ' Έκδοση)).

Αλγόριθμοι Τεχνητής Νοημοσύνης – Μηχανικής Μάθησης

Αλγόριθμος Random Forest

Ο αλγόριθμος του Τυχαίου Δάσους (Random Forest) είναι τύπος μηχανικής μάθησης με επίβλεψη (Supervised Learning - SL). Αναπτύχθηκε το 1995 από τον Tim Kam Ho χρησιμοποιώντας τη μέθοδο τυχαίου υποχώρου (random subspace method), και ήταν μία εφαρμογή της στοχαστικής διάκρισης (stochastic discrimination). Αργότερα, επεκτάθηκε και κατοχυρώθηκε με τον σημερινό τίτλο (Random Forest) από τον Leo Breiman και Adele Cutler, και εφαρμόζεται με το να κατασκευάζει συλλογή από δένδρα απόφασης με ελεγχόμενη διακύμανση (Forest, n.d.) (IBM_Random, n.d.).

Τα τυχαία δάση χρησιμοποιούνται συχνά ως μοντέλα μαύρου κουτιού καθώς δημιουργούν λογικές προβλέψεις, με την εφαρμογή τους, σε ένα ευρύ φάσμα δεδομένων και απαιτούν μικρή διαμόρφωση. Γενικά ο αλγόριθμος έχει πολλά δένδρα αποφάσεων ως βασικά μοντέλα εκμάθησης. Εκτελείται τυχαία δειγματοληψία σειρών και χαρακτηριστικών που σχηματίζουν δείγματα δεδομένων για κάθε μοντέλο. Αυτή το τμήμα του αλγορίθμου ονομάζεται Bagging.

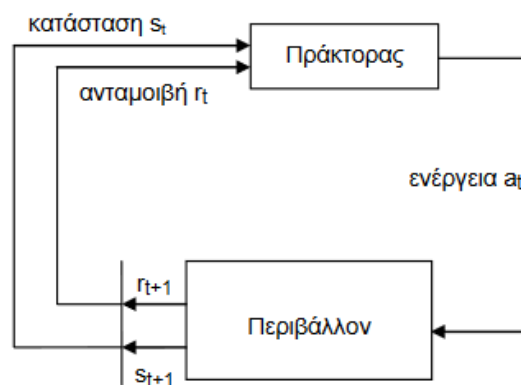
Ειδικότερα, ένα τυχαίο δείγμα δεδομένων επιλέγεται τυχαία για εκπαίδευση. Η επιλογή γίνεται με αντικατάσταση, το οποίο σημαίνει ότι μεμονωμένα σημεία των δεδομένων (data points) μπορούν να επιλεγούν περισσότερες από μία φορές. Εφόσον δημιουργηθούν πολλά δείγματα δεδομένων, αυτά τα μοντέλα εκπαιδεύονται ανεξάρτητα και ανάλογα με τον τύπο της εργασίας (παλινδρόμηση ή ταξινόμηση), ο μέσος όρος ή πλειοψηφία των προβλέψεων δίνει ακριβή εκτίμηση, με μειωμένη διακύμανση.

Ο αλγόριθμος του Τυχαίου Δάσους είναι μία επέκταση του Bagging, που οδηγεί στη δημιουργία ασύνδετων δένδρων αποφάσεων. Η τυχειότητα των χαρακτηριστικών, δημιουργεί ένα τυχαίο υποσύνολο χαρακτηριστικών, με χαμηλή συσχέτιση μεταξύ των δένδρων αποφάσεων. Ενώ στα Δένδρα Αποφάσεων λαμβάνονται υπόψιν όλες διαιρέσεις των χαρακτηριστικών, στο τυχαίο δάσος επιλέγεται μόνο ένα υποσύνολο αυτών των χαρακτηριστικών (IBM_Random, n.d.).

Ενισχυτική Μάθηση

Με τον όρο ενισχυτική μάθηση (Reinforcement Learning - RL) αναφερόμαστε σε μία οικογένεια τεχνικών με τις οποίες το σύστημα μάθησης προσπαθεί να μάθει μέσα από την άμεση αλληλεπίδραση με το περιβάλλον. Εφαρμόζεται στον έλεγχο κίνησης ρομπότ, στη μάθηση επιτραπέζιων παιχνιδιών κτλ. Η έννοια της ενισχυτικής μάθησης είναι εμπνευσμένη από τα αντίστοιχα ανάλογα της μάθησης με επιβράβευση και τιμωρία που συναντώνται ως μοντέλα των έμβιων όντων. Σκοπός είναι το σύστημα μάθησης να μεγιστοποιήσει μια συνάρτηση του αριθμητικού σήματος ενίσχυσης (ανταμοιβή). Το σύστημα δε καθοδηγείται από κάποιον εξωτερικό επιβλέποντα για το ποια ενέργεια θα πρέπει να ακολουθήσει, αλλά ανακαλύπτει μόνο του ποιες ενέργειες είναι αυτές που αποφέρουν το μεγαλύτερο κέρδος (Βλαχάβας Ι., 2011 (Γ' Έκδοση)).

Η πλέον διαδεδομένη διαδικασία μέσω της οποίας επιτυγχάνεται η ενισχυτική μάθηση είναι της δοκιμής και αποτυχίας (trial-and-error). Στο βασικό πλαίσιο δομής της ενισχυτικής μάθησης, βρίσκονται οι έννοιες του πράκτορα (agent) και της ανταμοιβής (rewards). Πράκτορας είναι η οντότητα που μαθαίνει κι παίρνει αποφάσεις. Οτιδήποτε άλλο εκτός του πράκτορα ονομάζεται περιβάλλον. Ο πράκτορας και το περιβάλλον αλληλοεπιδρούν συνεχώς με τον πρώτο να επιλέγει τις ενέργειες και το δεύτερο να αποκρίνεται σε αυτές και ντου παρουσιάζει νέες καταστάσεις. Το περιβάλλον δίνει στον πράκτορα ανταμοιβές (rewards), ειδικές αριθμητικές τιμές τις οποίες αυτός προσπαθεί μακροπρόθεσμα να τις μεγιστοποιήσει (Βλαχάβας Ι., 2011 (Γ' Έκδοση)).



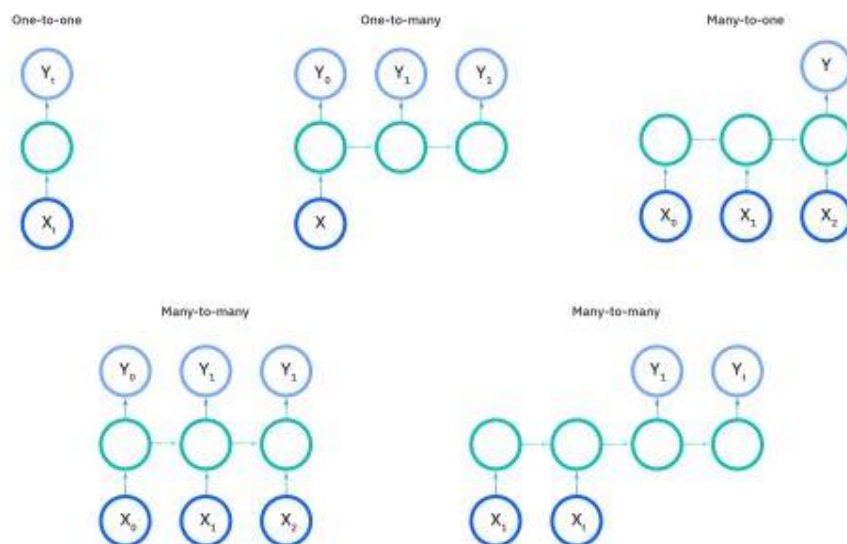
Εικόνα 9 Το πλαίσιο της ενισχυτικής μάθησης.

Τύποι Νευρωνικών Δικτύων – Μηχανικής Μάθησης

Αναδρομικό νευρωνικό δίκτυο

Το αναδρομικό νευρωνικό δίκτυο (Recurrent Neural Network – RNN) είναι τύπος τεχνητού νευρωνικού δικτύου, το οποίο χρησιμοποιεί διαδοχικά δεδομένα ή δεδομένα χρονοσειρών (time series). Είναι αλγόριθμος βαθιάς μάθησης και χρησιμοποιείται για τακτικά ή χρονικά προβλήματα όπως η μετάφραση της γλώσσας, η αναγνώριση ομιλίας, φωνητική αναζήτηση κλπ. (IBM_RNN, n.d.)

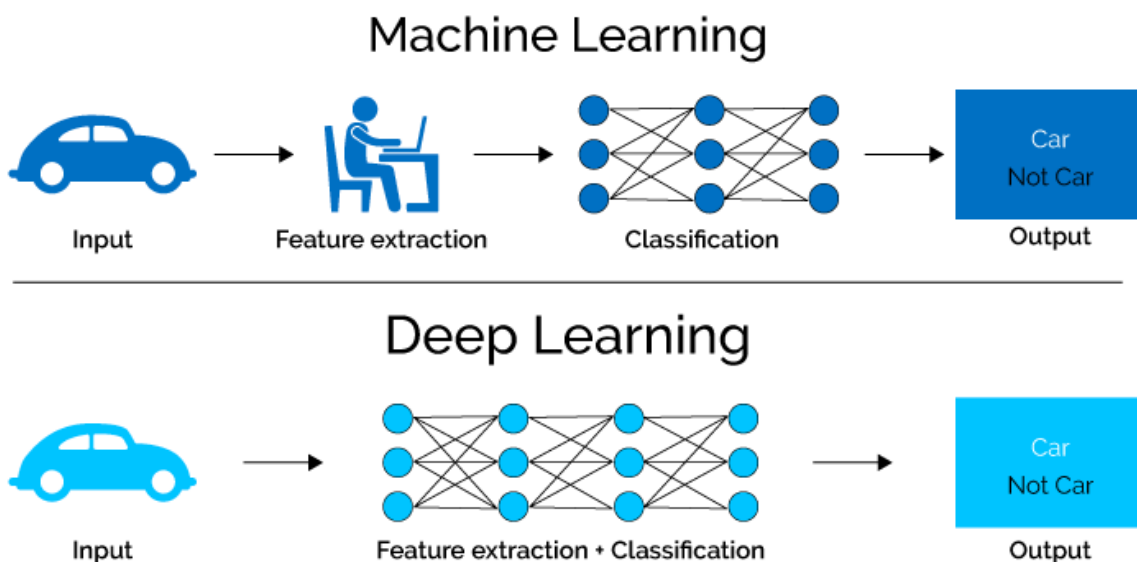
Τα αναδρομικά νευρωνικά δίκτυα διακρίνονται για τη «μνήμη» τους καθώς λαμβάνουν πληροφορίες από προηγούμενες εισόδους για να επηρεάσουν την τρέχουσα είσοδο και έξοδο, το οποίο οδηγεί στο φαινόμενο, η έξοδος τους να εξαρτάται από τα προηγούμενα στοιχεία της ακολουθίας, σε αντίθεση με άλλα (νευρωνικά δίκτυα) όπου είσοδος και έξοδος είναι ανεξάρτητες. Επίσης, χαρακτηριστικό τους είναι ότι ανάμεσα στα επίπεδα γίνεται μεταφορά των παραμέτρων, έτσι ενώ κάθε κόμβος έχει διαφορετικό συντελεστή «βάρους», αυτά μοιράζονται την ίδια παράμετρο βάρους σε κάθε επίπεδο του δικτύου. Αυτό, βοηθάει στην ενισχυτική μάθηση. Ο αλγόριθμος που αξιοποιούν είναι ο Backpropagation (IBM_RNN, n.d.).



Εικόνα 10 Τοπολογία RNN.

Βαθύ νευρωνικό δίκτυο

Το βαθύ νευρωνικό δίκτυο αποτελεί κλάδο της μηχανικής μάθησης. Ένα βαθύ νευρωνικό δίκτυο έχει ως σκοπό να «μάθει» μέσα από την άμεση παροχή της πληροφορίας σκοπού, την οποία θα κληθούν να εξασκηθούν και να μάθουν. Μέσα από την ανάπτυξη επιπέδων (layers) νευρωνικών συνδέσεων, καταφέρνει το σύστημα να μάθει τη πληροφορία που του δίνεται, δηλαδή τα εσωτερικά επίπεδα να έχουν κατάλληλη σύνδεση και βάρη, ώστε να αντιλαμβάνονται τη πληροφορία (είδος) που τους δίνεται. Για να λειτουργήσουν χρειάζονται, μεγάλο όγκο δεδομένων και αντίστοιχη υπολογιστική ισχύ. Χαρακτηριστικό παράδειγμα βαθύως νευρωνικού δικτύου είναι για την μελέτη και αναγνώριση εικόνων όπως τα Συνελεκτικά Νευρωνικά Δίκτυα, για την μελέτη και αναγνώριση εικόνων (IBM_DeepLearning, n.d.) (Semiengineering_DL, n.d.).



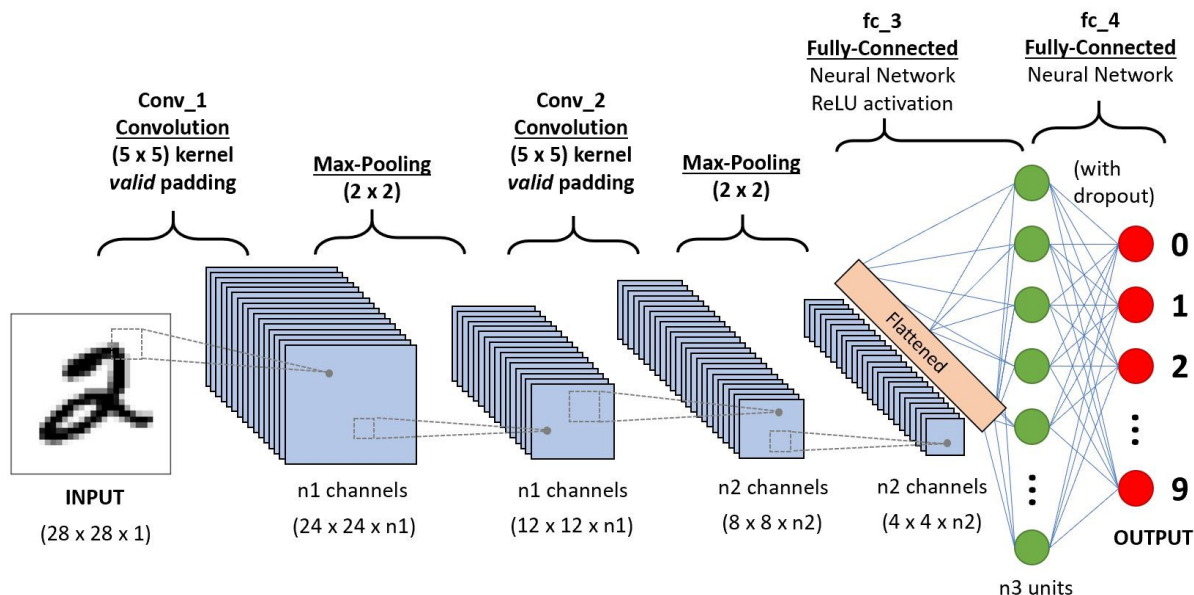
Εικόνα 11 Διαφορά μηχανικής μάθησης και Βαθιάς Μάθησης.

Συνελεκτικό Νευρωνικό Δίκτυο

Το συνελεκτικό νευρωνικό δίκτυο (Convolutional Neural Network) Πρόκειται για είδος βαθύς νευρωνικού δικτύου, τα οποία χρησιμοποιούνται κυρίως για την ανάλυση/αναγνώριση/μελέτη εικόνων και βίντεο.

Ένα συνελεκτικό νευρωνικό δίκτυο είναι μια κανονικοποιημένη μορφή ενός multilayer perceptron (MLP), όπου perceptron είναι αλγόριθμος υποβοηθούμενης μάθησης δυαδικών δεδομένων. Multilayer συνήθως σημαίνει πλήρως συνδεδεμένοι οι νευρώνες του ενός επιπέδου, με τους νευρώνες του επομένου. Αδυναμία μίας τέτοιας τοπολογίας αποτελεί το overfitting, δηλαδή της εξαγωγής μοντέλου, μόνο για τα δεδομένα εισαγωγής, και όχι με ικανότητα γενίκευσης δεδομένων ίδιας κατηγορίας πχ εικόνες. Για να γίνει αυτό τα CNN υιοθετούν, κατά την κανονικοποίηση, ιεραρχικά μοτίβα στα δεδομένα και συναρμολογούν μοτίβα αυξανόμενης πολυπλοκότητας, χρησιμοποιώντας μικρότερα και απλούστερα μοτίβα μέσω φίλτρων. Με αποτέλεσμα η κλίμακα συνδεδιμότητας και πολυπλοκότητας να είναι στη κατώτερη τιμή (CNN_wiki, n.d.) (CNN_analyticsvidhya, n.d.).

Βασικά επίπεδα των CNN είναι : το επίπεδο συνέλεξης, το επίπεδο αναμονής (pooling) και το πλήρους-σύνδεσης (Fully-Connected FC).



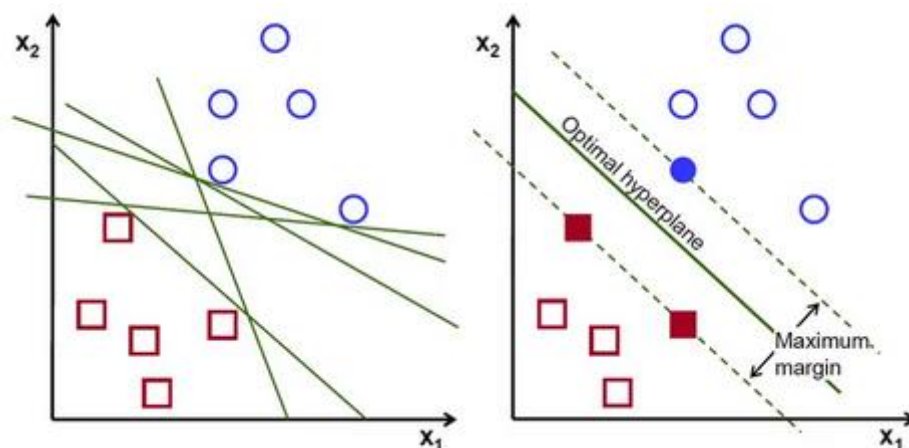
Εικόνα 12 Δομή CNN

Μηχανές διανυσμάτων υποστήριξης

Οι μηχανές διανυσμάτων υποστήριξης (Support Vector Machines – SVM), δημιουργήθηκαν από τον Vladimir Vapnik et al. το 1992 και στηρίζονται στην Θεωρία Στατιστικής Μάθησης (Statistical Learning Theory -SLT) και στα νευρωνικά δίκτυα τύπου Perceptron. Ως μέθοδοι (γραμμικής ή μη) παρεμβολής και ταξινόμησης, βρίσκουν εφαρμογή στην αναγνώριση γραφής (handwriting recognition), ταξινόμηση κειμένων (text recognition) και τη ταξινόμηση δεδομένων έκφρασης γονιδίων (gene expression data) (Βλαχάβας Ι., 2011 (Γ' Έκδοση)).

Στην εφαρμογή τους σε σενάριο ταξινόμησης, προσπαθούν να βρουν μια υπερεπιφάνεια (hypersurface) που να διαχωρίζει το χώρο των παραδειγμάτων (σε αρνητικά και θετικά). Η απόσταση της υπερεπιφάνειας από τα κοντινότερα θετικά και αρνητικά παραδείγματα, πρέπει να είναι η μέγιστη (maximum margin hypersurface) (Βλαχάβας Ι., 2011 (Γ' Έκδοση)).

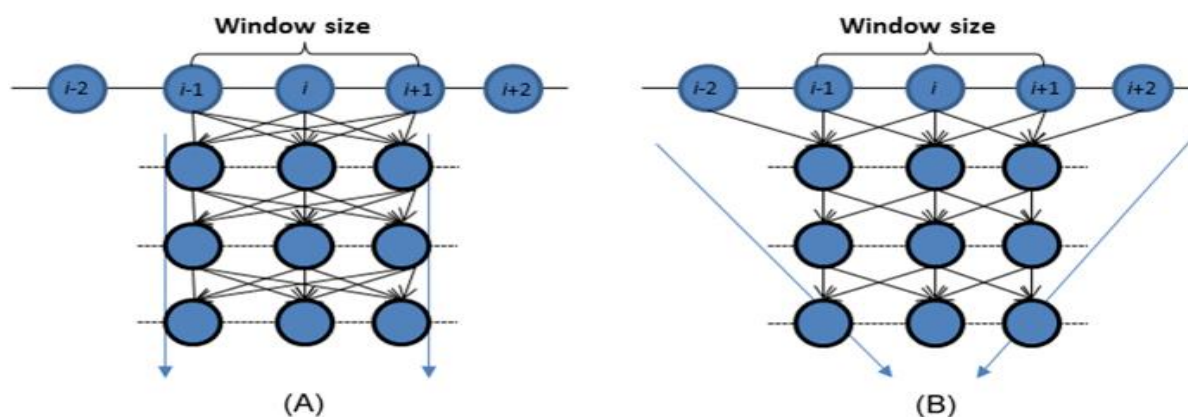
Το βασικό πλεονέκτημα των ΜΔΥ σε σχέση με τα νευρωνικά δίκτυα τύπου Perceptron είναι η παραγωγή πιο σύνθετων επιφανειών, έχοντας την ικανότητα ενσωμάτωσης μετασχηματισμών και συνδυασμών των αρχικών μεταβλητών ανάλογα με το πρόβλημα. Για το σκοπό αυτό χρησιμοποιούν ένα πεπερασμένο αριθμό υποσυνόλων του συνόλου εκπαίδευσης, που ονομάζονται διανύσματα υποστήριξης (support vectors) καθώς και συναρτήσεις πυρήνα (kernel functions), προκειμένου να μετασχηματίσουν τον αρχικό χώρο υποθέσεων και να βρουν τη βέλτιστη μη γραμμική υπερεπιφάνεια, με ελάχιστο σφάλμα ταξινόμησης (Βλαχάβας Ι., 2011 (Γ' Έκδοση)). (SVM_towardsdatascience, n.d.)



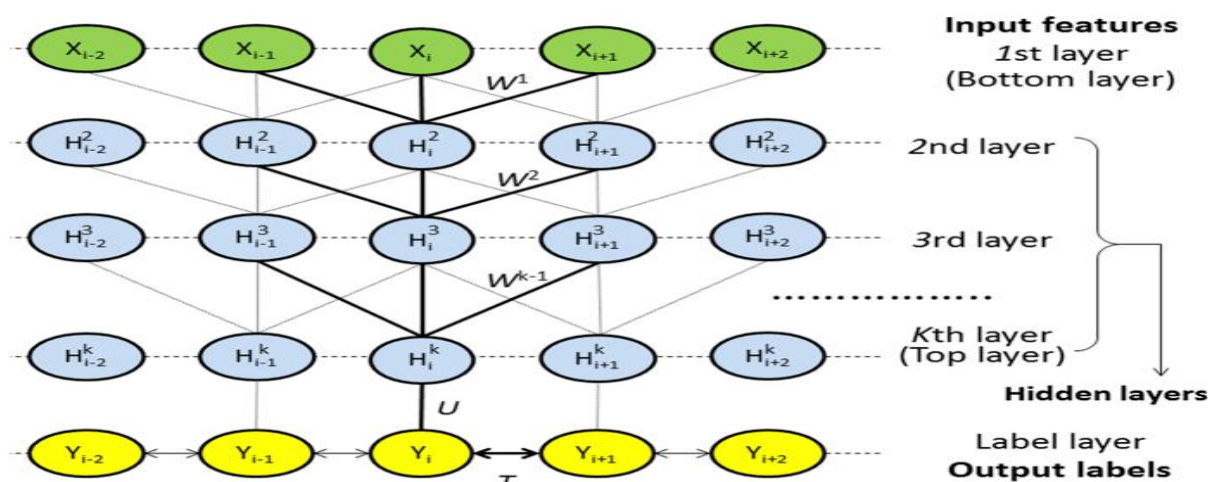
Εικόνα 13 Εφαρμογή ΜΔΥ

Deep Convolutional Neural Fields

Τα Βαθιά Συνελεκτικά Νευρωνικά Πεδία (Deep Convolutional Neural Fields – DCNF) αποτελούν μέθοδο μηχανικής μάθησης και αναπτύχθηκαν για την πρόβλεψη στις δευτεροταγούς δομής πρωτεϊνών. Τα DCNF συνδυάζουν τα πλεονεκτήματα και των Υπό Συνθήκη Νευρωνικών Πεδίων (Conditional Neural Fields – CNF) και των Βαθιών Συνελεκτικών Νευρωνικών Δικτύων (Convolutional Neural Networks – DCNN). Ο συνδυασμός αυτός παρέχει την ανάλυση και μελέτη όχι μόνο σύνθετων σχέσεων μεταξύ αλληλουχίας αμινοξέων και δομής πρωτεΐνης αλλά και να μοντελοποιούν τη συσχέτιση δομών μεταξύ γειτονικών περιοχών. Τα DeepCNF είναι όμοια με τα Υπο συνθήκη Τυχαία Πεδία (Conditional Random Fields – CRF) και CNF, στην μοντελοποίηση των πρωτεϊνών. Ωστόσο, τα DeepCNF χρησιμοποιούν DCNN για να το ρηχό βάθος νευρωνικού δικτύου που χρησιμοποιούν τα CNF, και να μπορεί να αποτυπώσει πολυπλοκότερες σχέσεις μεταξύ των χαρακτηριστικών εισόδου και εξόδου. Επίσης από το DCNN μπορεί να χρησιμοποιηθεί πληροφορία για μεγάλου μήκους αλληλουχίες (Εικόνα 14, 15). (Wang Sheng, 2016)



Εικόνα 14 Τυπικό CNF (A) και DeepCNF (B).



Εικόνα 15 Αρχιτεκτονική του DeepCNF

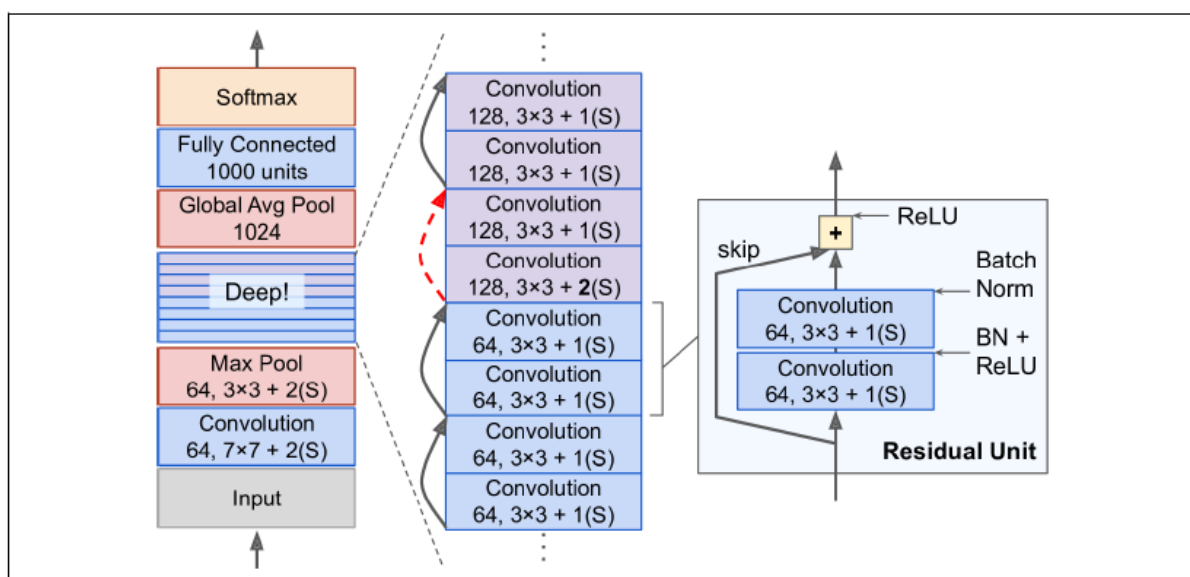
ResNet τοπολογία

Ο όρος ResNet προέρχεται από το Residual Network και είναι ένα τύπος συνελκτικού νευρωνικού δικτύου (CNN) και δημιουργήθηκε το 2015 από την ομάδα Kaiming et al. Πρόκειται για είδος CNN που χρησιμοποιούνται συνήθως, για να ενισχύσουν, εφαρμογές υπολογιστικής όρασης (He K, 2015).

Ειδικότερα το ResNet-50 είναι ένα συνελκτικό νευρωνικό δίκτυο 50 επιπέδων, εκ των οποίων 48 συνελκτικά επίπεδα, ένα επίπεδο MaxPool και ένα average pool layer. Τα Residual νευρωνικά δίκτυα είναι τύπος τεχνητού νευρωνικού δικτύου που σχηματίζει δίκτυα μέσω στοίβας blocks.

Ως προς την αρχιτεκτονική του, ακολουθεί τη βασική δομή του ResNet-34, περιέχει περισσότερα συνελκτικά επίπεδα, χωρίς να αντιμετωπίζει το πρόβλημα της διαβάθμισης, μέσω της σύνδεσης συντομεύσεων. Η σύνδεση συντόμευσης παρακάμπτει ορισμένα επίπεδα, μετατρέποντας ένα πρωτεύον (regular) δίκτυο σε υπολειπόμενο (residual) δίκτυο. (ResNet_datagen.tech, n.d.)

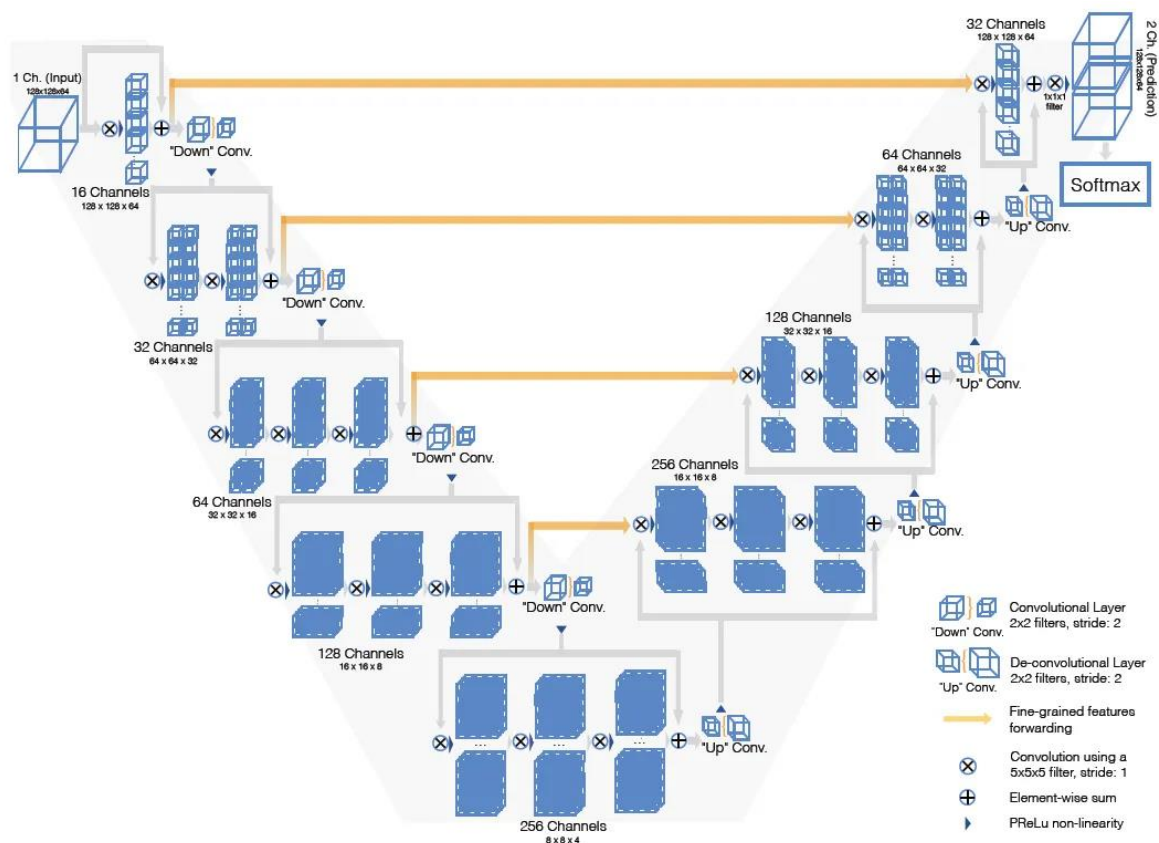
Η αρχιτεκτονική του ResNet-50 παρουσιάζεται στην παρακάτω εικόνα 11. Η softmax είναι συνάρτηση (softmax function), η οποία λαμβάνει ως είσοδο ένα διάνυσμα από K πραγματικούς αριθμούς και τον μετασχηματίζει σε πιθανότητες κατανομής (probability distribution) K πιθανών εξόδων. (Softmax_wikipedia, n.d.) (ResNet-50_jason-adam, n.d.)



Εικόνα 16 Αρχιτεκτονική του ResNet-50.

V-Net

Το V-Net είναι ένα συνελκτικό νευρωνικό δίκτυο (CNN), το οποίο αναπτύχθηκε για την ανάλυση και μελέτη τρισδιάστατων μαγνητικών εικόνων (3D images MRI). Αυτό αποτελείσε πρόκληση, λόγω των φυσικοανατομικών διαφορών που έχει ο ιστός στον τρισδιάστατο χώρο, από frame σε frame (Milletari F, 2016). Η αρχιτεκτονική του συγκεκριμένου μοντέλου φαίνεται στην εικόνα 17.



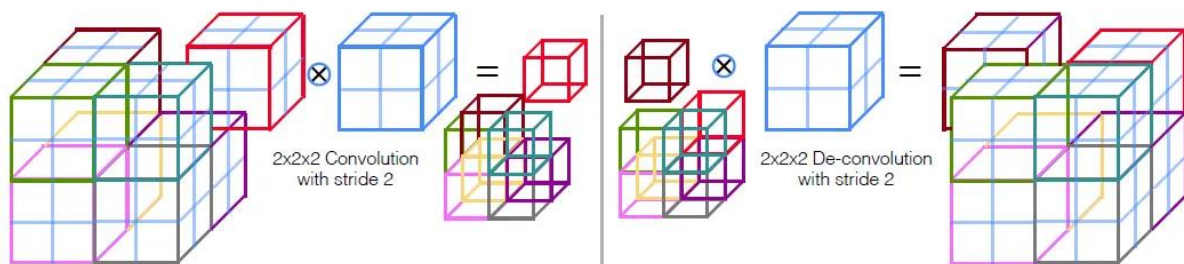
Εικόνα 17 Αρχιτεκτονική Vnet

Βασική μονάδα του μοντέλου είναι το voxel. Και χωρίζεται σε δυο τμήματα το αριστερό και το δεξί (Εικόνα 17). (Vnet_towardsdatascience, n.d.)

Στο αριστερό τμήμα είναι η λειτουργία μάθησης. Η είσοδος κάθε σταδίου χρησιμοποιείται από τα συνελκτικά επίπεδα και επεξεργάζεται μη γραμμικά (non-linearities) και προστίθεται στην έξοδο του τελευταίου συνελκτικού επιπέδου αυτού του σταδίου, ώστε να επάγει μάθηση στη διαφορική συνάρτηση (residual function). Αυτή τη αρχιτεκτονική εξασφαλίζει σύγκλιση σε σχέση με το δίκτυο μάθησης U-net. Οι συνελξεις που εκτελούνται σε κάθε στάδιο χρησιμοποιούν ογκομετρικούς πυρήνες μεγέθους 5x5x5 voxels (στοιχειώδης μονάδα όγκου).

Κατά τη συμπίεση, η ανάλυση μειώνεται με συνέλιξη πυρήνων πλάτους $2 \times 2 \times 2$ voxels. Με αποτέλεσμα το μέγεθος χαρακτηριστικών να μειώνεται στο μισό και παράλληλα ο αριθμός των καναλιών των χαρακτηριστικών να διπλασιάζεται σε κάθε στάδιο (της συμπίεσης).

Αντικαθιστώντας τη διαδικασία pooling με το βήμα convolutional ones, έχει ως αποτέλεσμα μειωμένη χρήση υπολογιστικής μνήμης κατά την εκπαίδευση του μοντέλου, διότι δεν υπάρχουν εναλλαγές (switches) από την έξοδο του pooling στις εισόδους για την αντίστροφη κατεύθυνση (back-propagation). (Vnet_towardsdatascience, n.d.)



Εικόνα 18 Μεθοδολογία Downsampling (αριστερά) και Upsampling (δεξιά).

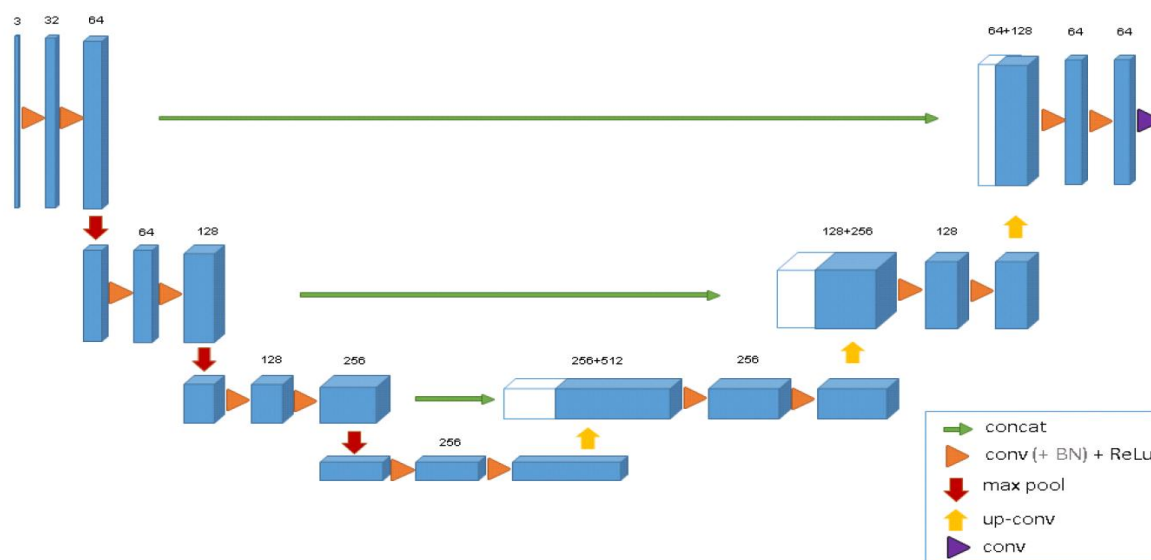
Στη δεξιά πλευρά του γραφήματος, έχουμε το στάδιο κατά το οποίο γίνεται η εξαγωγή των χαρακτηριστικών και ενίσχυση (expands) χωρικά των χαρτών (των χαρακτηριστικών) χαμηλής ανάλυσης, ώστε να συγκεντρώσει και να συνθέσει τις απαραίτητες πληροφορίες για την έξοδο σε δύο κανάλια, μιας ογκομετρικής τμηματοποίησης.

Σε κάθε στάδιο, εφαρμόζεται αποσυνέλιξη προκειμένου να αυξηθεί το μέγεθος των εισόδων ακολουθούμενα από ένα ως τρία συνελκτικά στρώματα, που περιλαμβάνουν το ήμισυ του αριθμού των πυρήνων $5 \times 5 \times 5$ που χρησιμοποιήθηκαν στο προηγούμενο επίπεδο. Η διαφορική συνάρτηση μαθαίνει παρόμοια με το αριστερό μέρος του δικτύου. Οι δυο χάρτες χαρακτηριστικών που υπολογίζονται από το τελευταίο συνελκτικό επίπεδο, έχουν μέγεθος πυρήνα (kernel) $1 \times 1 \times 1$ και παράγουν έξοδο ίδιου μεγέθους με τον όγκο εισόδου. Αυτοί οι δυο χάρτες χαρακτηριστικών στην έξοδο είναι οι πιθανότητες των τμηματοποιήσεων των περιοχών, και εφαρμόζονται στην soft-max voxelwise. (Vnet_towardsdatascience, n.d.)

Παράλληλα στην οριζόντια σύνδεση, παρόμοια με το U-net, οι πληροφορίες θέσεις χάνονται (is lost) στη συμπίεση, με αποτέλεσμα τα χαρακτηριστικά που εξάγονται από τα πρώτα στάδια του αριστερού τμήματος στο CNN να προωθούνται στο δεξί τμήμα με οριζόντια σύνδεση. Αυτή η μεταφορά μπορεί να βοηθήσει σχετικά με τη πληροφορία για τη θέση, καθώς και βελτίωση τη ποιότητας του τελικού περιγράμματος. Επίσης αυτές οι συνδέσεις βελτιώνουν το χρόνο σύγκλισης του μοντέλου.

3D U-net

Το 3D U-net είναι μέθοδος επεξεργασίας ογκομετρικών δεδομένων, με αποτέλεσμα να βρίσκει εφαρμογή στο τομέα των ιατρικών αναλύσεων. Βασίζεται στην αρχιτεκτονική του U-net, αποτελούμενο από έναν κωδικοποιητή για την ανάλυση ολόκληρης της εικόνας και ένα τμήμα αποκωδικοποιητή για την εξαγωγή τμήματος σε υψηλή ανάλυση (full resolution). Το 3D U-net λαμβάνει στην είσοδο τρισδιάστατα δεδομένα και εφαρμόζει, σε αντίθεση με το 2D U-net, 3D περιέλιξη (convolution), 3D macpooling και 3D up-convolution. (Çiçek Ö A. A., 2016)



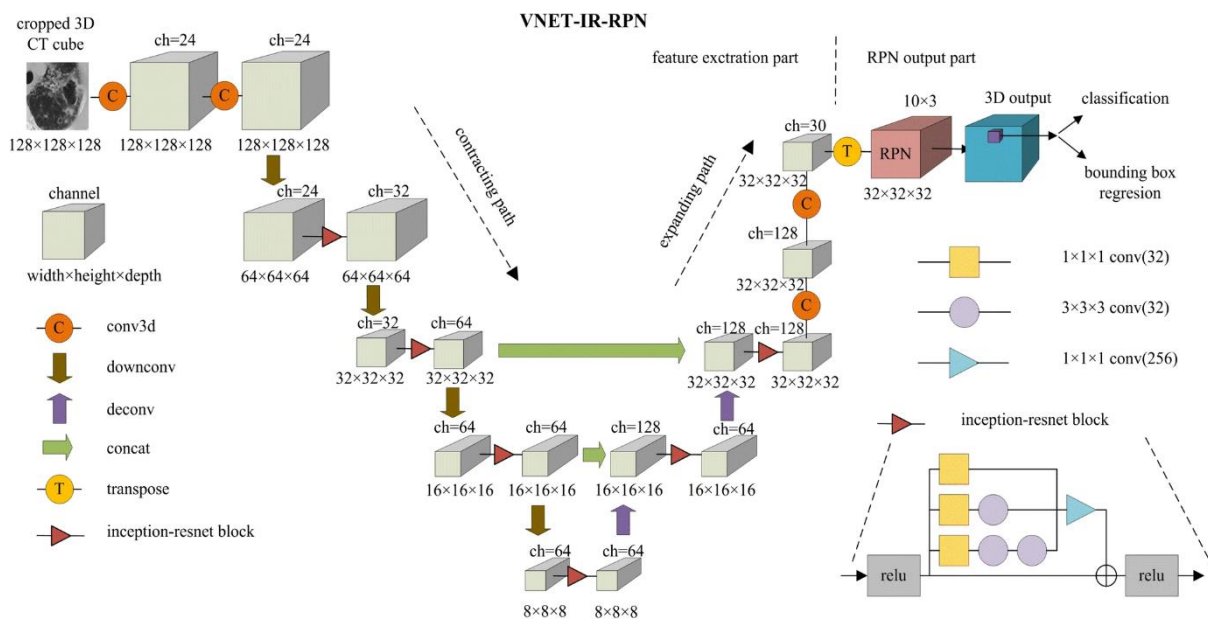
Εικόνα 19 Αρχιτεκτονική 3D U-Net.

Η αρχιτεκτονική του 3D U-net περιλαμβάνει δυο τμήματα, ένα τμήμα ανάλυσης (στα αριστερά) και ένα τμήμα σύνθεσης (στα δεξιά). Στο τμήμα της ανάλυσης υπάρχουν δυο συνελίξεις (convolutions) $3 \times 3 \times 3$ σε κάθε επίπεδο. Κάθε ένα από τα επίπεδα ακολουθείται από ένα ReLU (rectified linear unit) και στη συνέχεια maxpooling $2 \times 2 \times 2$ με βήμα 2 σε κάθε διάσταση. Στη φάση της σύνθεσης κάθε επίπεδο αποτελείται από μία upconvolution $2 \times 2 \times 2$ με βήμα 2 σε κάθε διάσταση. Ακολουθείται από δυο $3 \times 3 \times 3$ convolution, όπου η κάθε μία ένα ένα ReLU. Κατά τη φάση της σύνθεσης τα βασικά χαρακτηριστικά υψηλής ανάλυσης παρέχονται μέσω συντομεύσεων μεταξύ επιπέδων ίσης ανάλυσης (στη διαδρομή της ανάλυσης). Τέλος στην έξοδο εφαρμόζεται μία συνέλιξη $1 \times 1 \times 1$, η οποία μειώνει τον αριθμό των καναλιών στα 3. Στο σύνολό της η αρχιτεκτονική έχει 19069955 παραμέτρους. Για να αποφευχθούν τα σημεία συμφόρησης (bottleneck), ο αριθμός των καναλιών διπλασιάζεται πριν από το max-pooling. Κάθε μπλοκ χρησιμοποιεί κανονικοποίηση (normalization) πριν το ReLU. Κατά την εκπαίδευση κάθε μπλοκ κανονικοποιείται με τον μέσο όρο και την τυπική απόκλιση. (3D_U-Net_iq.opennngenus.org, n.d.)

VNet-inception-residual network

Το V-Net είναι μία 3D έκδοση του U-Net με resnet blocks. Αυτός ο συνδυασμός των δυο μοντέλων έχει ιδιαίτερα οφέλη για τη μελέτη 3D δομών, και αυτό διότι πέρα της ύπαρξης του του V-Net, μας παρέχει μία προσαρμοστικότητα/πλαστικότητα στον εντοπισμό, όταν τα όρια της δομής που θα αναγνωρισθεί δεν είναι στατικά σε όλα τα υποκείμενα προς ανάλυση, πχ ιατρικά δεδομένα.

Επομένως, γίνεται χρήση του inception-resnet για την εξαγωγή των χαρακτηριστικών χρησιμοποιώντας πυρήνες συνέλιξης (convolution kernel) διαφορετικού μεγέθους. Το block του inception-resnet περιλαμβάνει διαδικασίες (operations) με διαφορετικού μεγέθους πυρήνες (1x1x1, 3x3x3), επιτρέποντας στο μοντέλο να εξάγει λεπτομερέστερα τα χαρακτηριστικά από διάφορα επίπεδα λήψης, μειώνοντας σημαντικά τις παραμέτρους του δικτύου. Επιπλέον, η λειτουργία concatenate εντός του block έναρξης χρησιμοποιήθηκε για να αμαλαγματώνει πολλαπλής κλίμακας (multi-scale) χαρακτηριστικά, βελτιώνοντας την εκπαίδευση του μοντέλου στην εξαγωγή των χαρακτηριστικών (Xukun Li, 2021).



Εικόνα 20 Αρχιτεκτονική του VNET-IR-RPN

EfficientNet B4

Η EfficientNetB4 είναι μία συνελεγκτική αρχιτεκτονική νευρωνικού δικτύου, η οποία εφαρμόζει κλιμάκωση/επέκταση του νευρωνικού δικτύου, ομοιόμορφα και προς όλες τις διαστάσεις βάθους/πλάτους/ανάλυσης, με την εφαρμογή ενός σύνθετου συντελεστή. Σε αντίθεση με τη συμβατική πρακτική που κλιμακώνει αυθαίρετα αυτούς τους παράγοντες, η μέθοδος κλιμάκωσης EfficientNet κλιμακώνει με ένα σύνολο σταθερών συντελεστών κλιμάκωσης (Tan M, 2019).

Για παράδειγμα σε περίπτωση ανάλυσης εικόνων, που μπορούμε να χρησιμοποιήσουμε επιπλέον (έως και πολλαπλάσιους) υπολογιστικούς πόρους, τότε μπορούμε απλώς να αυξήσουμε το βάθος και το πλάτος και το μέγεθος της εικόνας, όπου οι σταθεροί συντελεστές καθορίζονται μέσω μίας αναζήτησης μικρού πλέγματος στο αρχικό μικρό μοντέλο.

Η μέθοδος σύνθεσης της κλίμακας βοηθάει στην επεξεργασία εικόνων όπου σωστή επέκταση του μοντέλου μπορεί να αναγνωρίζει πιο αδρά μοτίβα, σε μεγαλύτερης ανάλυσης εικόνες.

Στην αναγνώριση εικόνων το EfficientNet έχει επιτύχει ακρίβεια 91.7% στο CIFAR-100 και 98.8% στο Flowers.

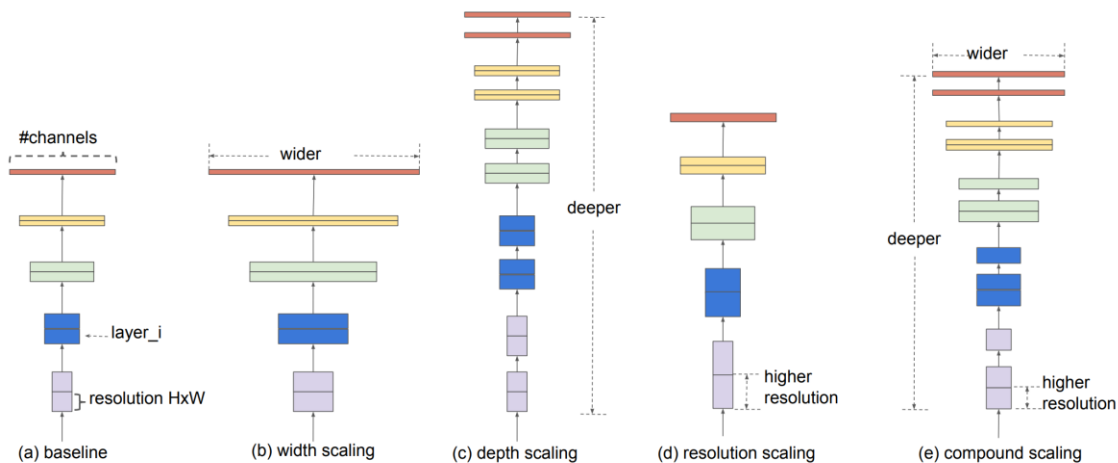


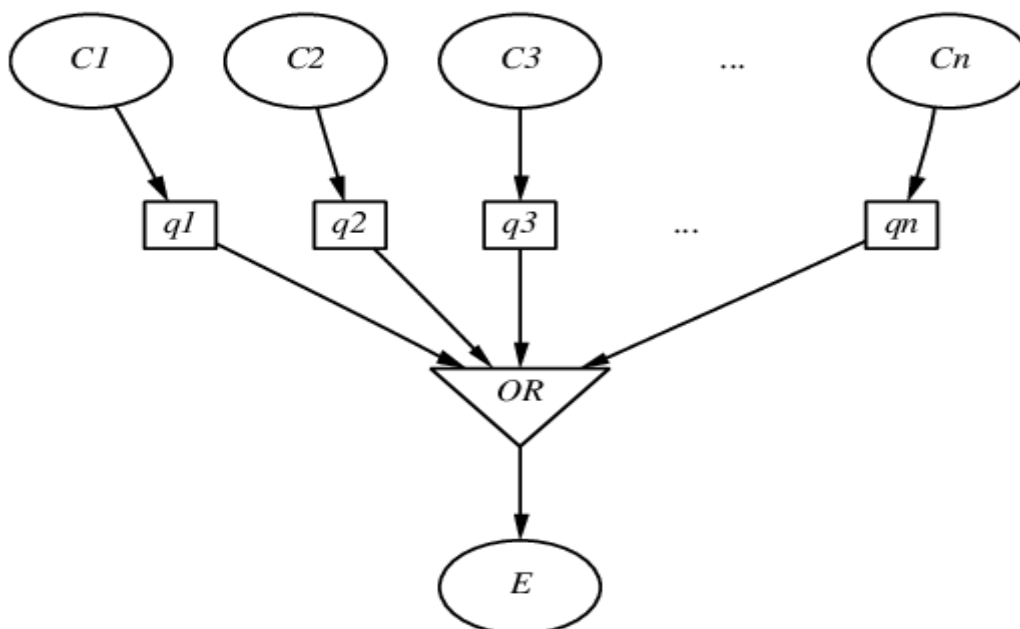
Figure 2. **Model Scaling.** (a) is a baseline network example; (b)-(d) are conventional scaling that only increases one dimension of network width, depth, or resolution. (e) is our proposed compound scaling method that uniformly scales all three dimensions with a fixed ratio.

Εικόνα 21 EfficientNet. κλιμάκωση.

Στατιστικές μέθοδοι επαλήθευσης/ανάλυσης

Στατιστική μέθοδο Noisy-OR Bayesian

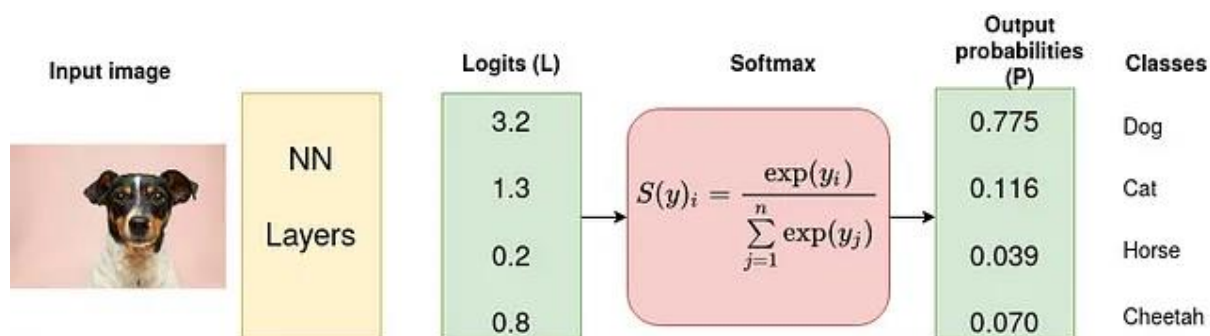
Ο μοντέλο δικτύου Noisy-OR (Noisy-OR model) είναι ειδική περίπτωση της ειδικεύσης των πινάκων πιθανότητας κατάστασης (conditional probability tables – CPT), στα δίκτυα Bayesian (Igi_global_noisy-or, n.d.). Το Bayesian δίκτυο, είναι μοντέλο γράφου, στο οποίο έχουμε τις πιθανότητες σύνδεσης/σχέσης μεταξύ των διαφόρων μεταβλητών και για την αναπαράσταση (αυτής της σχέσης) εφαρμόζονται μη κυκλικός κατευθυνόμενος γράφος (Directed Acyclic Graph -DAG), του οποίου τα βέλη αντιπροσωπεύουν τις αιτιατές επιρροές μεταξύ των κόμβων. Η ιδέα του Noisy-OR, είναι ότι σε ένα τέτοιο δίκτυο, κάθε κόμβος-γονέας αποδίδει/επηρεάζει τον επόμενο κόμβο-τέκνο, ανεξάρτητα από τους υπόλοιπους κόμβους-γονείς, και η επιμέρους επίδραση αναπαρίσταται μέσω της συνάρτησης Boolean OR (Εικόνα 9) (Korb, 2002) (Βλαχάβας Ι., 2011 (Γ' Έκδοση)). Σε ένα τέτοιο δίκτυο ο αριθμός των παραμέτρων είναι γραμμικός ως προς τον αριθμό των κόμβους-γονέων του κάθε κόμβου. Επίσης, σε ένα κλασικό σενάριο CPT η πολυπλοκότητα για μια δυαδική μεταβλητή με n απογόνους κόμβους (predecessors), σε ένα δίκτυο Bayes (Bayes Network) είναι 2^n . Με την εφαρμογή Noisy-OR, η πολυπλοκότητα γίνεται λογαριθμική. Χαρακτηριστική ιατροβιολογική εφαρμογή είναι σε μελέτη για το ήπαρ και το άσθμα, για υπολογισμό πρόγνωσης, και σχέσεις αιτιών/ασθένειας (Anand Vibha, 2008).



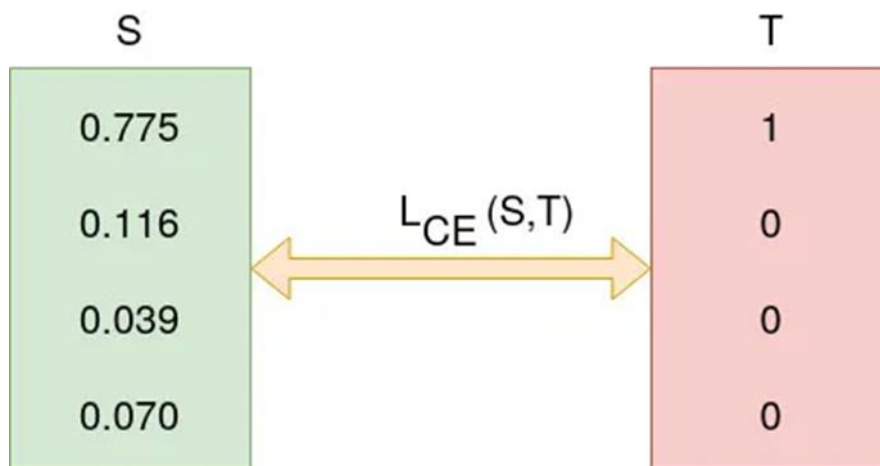
Εικόνα 22 Πύλη Noisy-OR Bayesian.

Διασταυρούμενη εντροπία

Πρόκειται για τη λεγόμενη cross entropy και αναφέρεται στην στατιστική ανάλυση, αξιολόγηση των αποτελεσμάτων ενός συστήματος μηχανικής μάθησης. Πρόκειται για μία λογαριθμική στατιστική μέτρηση, των αποτελεσμάτων του συστήματος και ειδικότερα η απόσταση που έχουν τα αποτελέσματα από τα αναμενόμενα (true) αποτελέσματα. Με την εφαρμογή της συνάρτησης Softmax έχουμε στατιστική απεικόνιση αυτής της απόκλισης. Όσο το σύστημα εκπαιδεύεται τόσο τα δεδομένα εξόδου θα συγκλίνουν με τα αναμενόμενα. (Cross_entropy, n.d.)



Εικόνα 23 Διασταυρούμενη Εντροπία και Συνάρτηση Softmax.



Εικόνα 24 Αποτελέσματα σε πιθανότητες και τα αναμενόμενα αποτελέσματα.

Wald test

Η στατιστική μέθοδος Wald (Wald test, Wald Chi-Squared Test) είναι ένα στατιστικό εργαλείο, το οποίο μας δίνει τη σημαντικότητα μίας μεταβλητής/παραμέτρου, στο σύστημά μας. Σημαντικότητα, δηλαδή κατά πόσο η μεταβλητή/παραμέτρος είναι απαραίτητη (ή πόσο επηρεάζει) το σύστημά μας, τόσο στη λειτουργία του, όσο και στη περιγραφή του (statisticshowto, n.d.). (Wald_test, n.d.)

Ο μαθηματικός τύπος της μεθόδου:

$$W_T = \frac{[\hat{\theta} - \theta_0]^2}{1/I_n(\hat{\theta})} = I_n(\hat{\theta}) [\hat{\theta} - \theta_0]^2$$

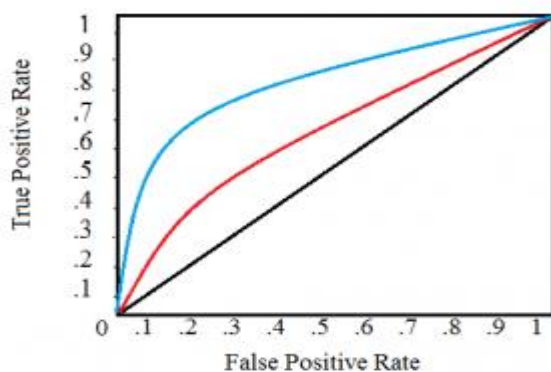
Εικόνα 25 Μαθητικός τύπος Wald test.

Όπου: Θ = Εκτιμητής Μέγιστης Πιθανότητας (Maximum Likelihood Estimator - MLE),

$I_n(\Theta)$ = αναμενόμενη Fisher πληροφορία (εκτιμήθηκε στην MLE).

C-Index

Η C-statistic ή όπως καλείται C-index, είναι μία στατιστική μέθοδος, η οποία μετρά την προσαρμογή (goodness of fit) δυαδικών αποτελεσμάτων, σε ένα μοντέλο παλινδρόμησης. Η μέθοδος αυτή είναι πολύ χρήσιμη σε κλινικές μελέτες, όπου μας δίνει τη πιθανότητα ένας τυχαία επιλεγμένος ασθενής που βίωσε ένα συμβάν (πχ μία ασθένεια ή κατάσταση) έχει υψηλότερο βαθμό κινδύνου από έναν ασθενή που δεν είχε βιώσει το συμβάν. Η C-index είναι ίση με τη περιοχή κάτω από τη καμπύλη Χαρακτηριστικού Δείκτη Λειτουργίας (Receiver Operating Characteristic – ROC) και έχει τιμή από 0 έως 1 (C-index, n.d.).



Εικόνα 26 Η στατιστική περιοχή είναι ίση με περιοχή κάτω από τη καμπύλη ROC.

η τιμή κάτω από 0.5 είναι αδύναμο μοντέλο
τιμή 0.5 το μοντέλο δε διαφέρει από ένα τυχαία επιλογή
τιμή άνω του 0.7 καλό μοντέλο
τιμή άνω του 0.8 δυνατό (strong) μοντέλο
τιμή 1 το μοντέλο ταιριάζει τέλεια στην ομάδα μελέτη (ασθενών και μη ασθενών).

AMSGrad

Το AMSGrad είναι στοχαστική μέθοδος βελτίωσης που επιδιώκει να διορθώσει ένα πρόβλημα σύγκλισης μέσω βελτιστοποιητών, οι οποίοι βασίζονται στη μέθοδο Adam. Το AMSGrad χρησιμοποιεί το μέγιστο των τετραγώνων των προηγούμενων διαβαθμίσεων u_t , αντί του εκθετικού μέσου όρου για να ενημερώσει τις παραμέτρους. (Reddi J. Sashank, 2018) (AMSGrad, n.d.)

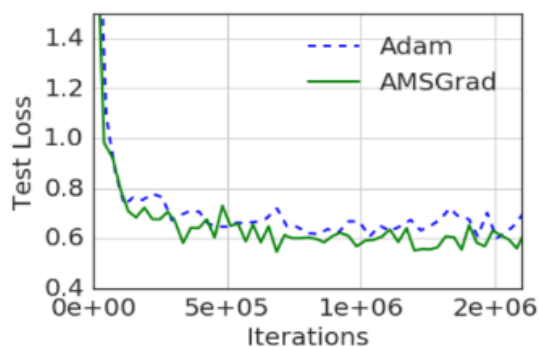
$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$$

$$\hat{v}_t = \max(\hat{v}_{t-1}, v_t)$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t + \epsilon}} m_t$$

Εικόνα 28 Μαθηματικός τύπος Wald test.



Εικόνα 27 Σύγκριση μεθόδων.

Adam

Ο Adam είναι αλγόριθμος βελτιστοποίησης στοχαστικών συναρτήσεων με βάση το διαφορικό πρώτης τάξης, σύμφωνα με τις εκτιμήσεις ροπών χαμηλότερης τάξης. Η μέθοδος είναι υπολογιστικά αποδοτική, έχει ελάχιστες απαιτήσεις σε μνήμη, και είναι κατάλληλη για προβλήματα που έχουν μεγάλα δεδομένα και παραμέτρους. Επίσης είναι κατάλληλη για προβλήματα με θόρυβο ή μικρά διαφορικά. Γενικότερα είναι ένας προσαρμοστικός αλγόριθμος βελτιστοποίησης ρυθμού εκμάθησης και χρησιμοποιείται τόσο για τη στιγμή όσο και σε κλιμάκωση, συνδυάζοντας τα οφέλη του RMSProp και του SGD (Adam, n.d.).

$$w_t = w_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}}$$

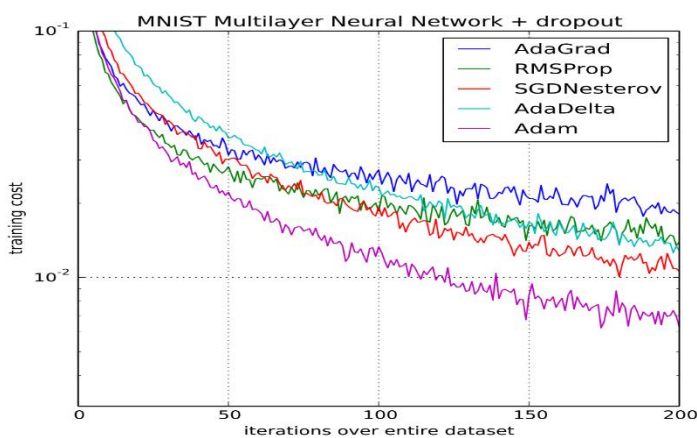
$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$$

Εικόνα 30 Μαθηματικός τύπος Adam.



Εικόνα 29 Σύγκριση μεθόδων.

Κύριο Μέρος

Η Τεχνητή Νοημοσύνη στον covid-19

Εφαρμογές σε CT-X ακτινογραφίες

Μοντέλο ομάδας Xu et al.

Το 2019 η ομάδα Xu et al. (Xu X., 2020) δημοσίευσαν μοντέλο διάγνωσης για τον Covid-19, μέσω της επεξεργασίας εικόνων από Υπολογιστικής Ακτινογραφίας (CT scan). Το μοντέλο αυτό είναι επιπέδου νευρωνικού δικτύου.

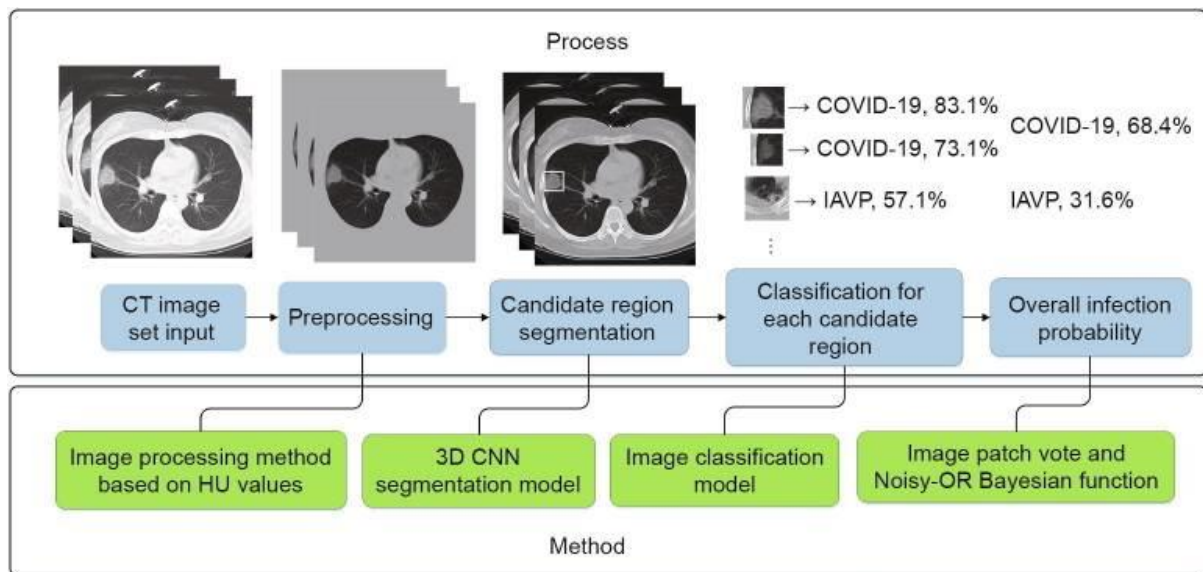
Δεδομένα/γλικά

Συνολικά χρησιμοποιήθηκαν στην μελέτη 618 transverse-section CT scan, από τις οποίες 219 ήσαν από 110 ασθενείς με Covid-19, των νοσοκομείων First Affiliated Hospital, College of Medicine, Zhejiang University, από το Wenzhou Central Hospital και από το First People's Hospital of Wenling, το χρονικό διάστημα από 19 Ιανουαρίου 2020, έως 14 Φεβρουαρίου 2020. Όλοι οι ασθενείς είχαν ταυτοποιηθεί διαγνωστικά με PCR-test covid-19.

Παράλληλα, οι υπόλοιπες 399 ακτινογραφίες ήσαν από το First Affiliated Hospital, College of Medicine, Zhejiang University και ήσαν ομάδα ελέγχου (control). Ειδικότερα 224 CT από ασθενείς με IAVP (εμπεριέχοντας H1N1, H3N2, H5N1, H7N9), με τη ταυτοποίησή τους να γίνεται με RT-PCR, και 175 δείγματα CT από υγιείς ανθρώπους (Xu X., 2020).

Η διαδικασία επεξεργασίας των CT image

Συνολικά υπάρχουν 3 στάδια συλλογής/επεξεργασίας των εικόνων (Εικόνα 25). Στο πρώτο στάδιο έχουμε τον καθορισμό των πνευμονικών περιοχών. Δεύτερον τη δημιουργία του 3D-μοντέλου, ώστε να έχουμε τη σωστή τμηματοποίηση των εικόνων σε cubes/υποπεριοχές των εικόνων. Τρίτο στάδιο, η χρήση μοντέλου διαχωρισμού/ταξινόμησης σε τρεις κατηγορίες COVID-19, IAVP, ITI (irrelevant to infection). Τμήματα εικόνων που είχαν «σημάδι» στο ίδιο cube, «σημειώνονταν» βαθμολογικά για την εγκυρότητά της, ως υποψήφια στο σύνολο της δεξαμενής εικόνων. Η όλη ανάλυση έγινε με χρήση τη στατιστική μέθοδο Noisy-OR Bayesian (Onísko A., 2001).



Εικόνα 31 Στάδια επεξεργασίας εικόνων.

Δημιουργία μοντέλου μάθησης

Έχοντας τη δεξαμενή δεδομένων (εικόνες CT) έγινε η προσπάθεια σύνταξης μίας νέας μεθόδου βαθειάς μάθησης και νευρωνικού δικτύου για την αυτοματοποιημένη διάγνωση.

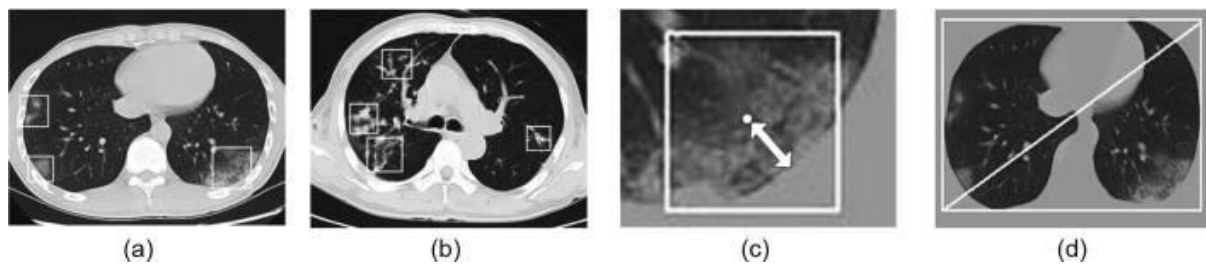
Η προσπάθεια στηρίχθηκε σε υφιστάμενες προσπάθειες/μοντέλα επεξεργασίας δεδομένων, και ειδικότερα σε μελέτη πνευμονικής φυματίωσης.

Ειδικότερα, ως βάση, έχουμε το μοντέλο VNet-inception-residual network (IR)-region proposal network (RPN) (Wu W, 2019), το οποίο χρησιμοποιεί το VNet (Milletari F, 2016) με βάση μία IR δομή (Szegedy C, 2016) ως επέκταση και στη συνέχεια ακολουθείται από RPN, για να εντοπίσει υποψήφιες περιοχές.

Στη 3D segmentation, δεν εφαρμόστηκε η εφαρμοζόμενη pixel διαχωρισμός όπως το VNet ή το 3D u-net (Çiçek Ö A. A., 2016), διότι η φύση του ιστού και της λοίμωξης, δεν επιτρέπουν τον σαφή διαχωρισμό των αλλοιώσεων. Αντί αυτού εφαρμόστηκε RPN, ως άλλος αλγόριθμος ανίχνευσης εικόνας/αντικειμένων, όπου έχουμε τρισδιάστατα πλαίσια οριοθέτησης αντί για τμηματοποίηση της μάσκας σε pixel.

Εντοπισμός περιοχής

Το κρίσιμο σημείο για την ανάπτυξη της μεθόδου είναι τα σημεία στις εικόνες όπου πρέπει να εστιάσει ο αλγόριθμος. Βασισμένοι σε μελέτες, παρατηρήθηκαν/διαπιστώθηκαν οι περιοχές ενδιαφέροντος για τη διάγνωση, είναι στην ground-glass περιοχή, στη περιφέρεια αλλοιώσεις κατά μήκος της περιοχής του υπεζυκώτα και σε περιπτώσεις ανεξάρτητων εστιών μόλυνσης. (Εικόνα 31).



Εικόνα 32 Περιοχές εστιών μόλυνσης (a) Covid-19 ground-glass , (b) IAVP εστίες, (c) ελάχιστη απόσταση μάσκας κέντρου φλεγμονής, (d) Διαγώνιος ελάχιστου περιγεγραμμένου ορθογωνίου

Το μοντέλο βασίζεται σε αυτά τα ευρήματα για τις βελτιστοποιήσεις του. Επίσης η ταξινόμηση των εικόνων έγινε στόχος για το μοντέλο, ώστε να διακρίνει αυτές τις δομές. Τέλος, εφαρμόστηκε, ως επιπλέον βαθμός «συντελεστής» σημαντικότητας, η απόσταση από τα όρια του πνεύμονα. Ο αλγόριθμος που εφαρμόστηκε για την απόσταση ήταν:

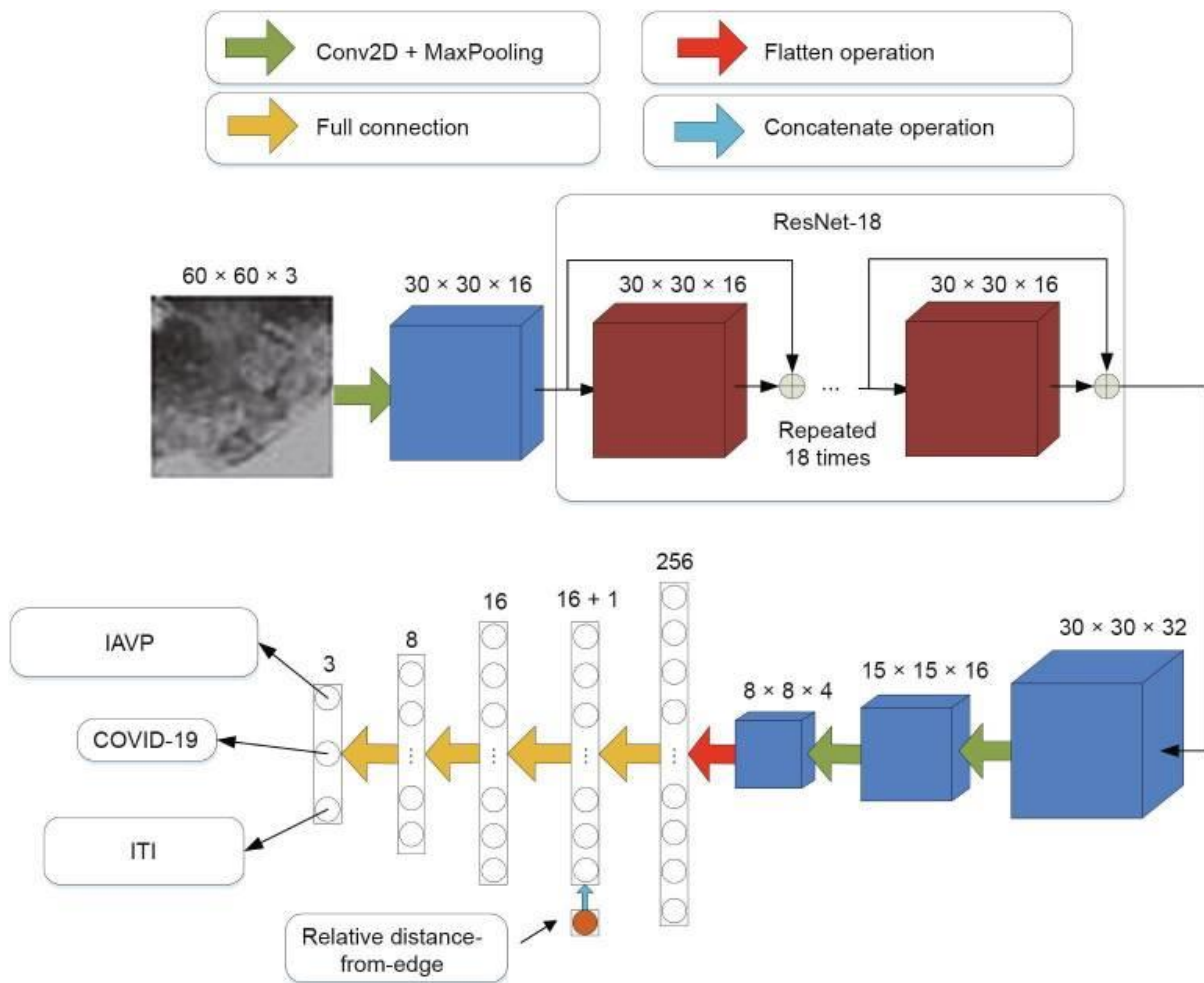
Βήμα 1: Υπολογισμός ελάχιστης απόστασης της φλεγμονής/σημείου (pach) από το κέντρο.

Βήμα 2: Υπολογισμός διαγωνίου του ελάχιστου περιγεγραμμένου ορθογωνίου.

Βήμα 3: Υπολογισμός σχετικής απόστασης από την άκρη (edge) διαιρώντας την απόσταση από το βήμα 1 με τη διαγώνιο από το βήμα 2.

Δομή Νευρωνικού Δικτύου

Δύο μοντέλα ταξινόμησης αναπτύχθηκαν/εφαρμόστηκαν στο συγκεκριμένο μοντέλο. (Εικόνα 32). Το ένα ήταν ένα παραδοσιακό παραμονής δίκτυο (residual network ResNet based) μοντέλο (He K, 2015) και το άλλο σχεδιάστηκε πάνω στη δομή του πρώτου δικτύου με εστίαση σε τοπικό επίπεδο σε σύνδεση με το πλήρες επίπεδο (full-connection layer), ώστε να βελτιώνει την ακρίβεια. Αυτός ο μηχανισμός προστέθηκε στο πρώτο επίπεδο πλήρους σύνδεσης (first full-connection layer) για να μεγιστοποιήσει την επίδραση αυτού του παράγοντα σε όλο το δίκτυο.



Εικόνα 33 Σχηματική αναπαράσταση νευρωνικού δικτύου

Η κλασική δομή του δικτύου Res-Net χρησιμοποιήθηκε για την εξαγωγή των χαρακτηριστικών των εικόνων. Επίσης μία δεξαμενή λειτουργιών (pooling operations) εφαρμόστηκαν ώστε να μειωθούν οι διαστάσεις των δεδομένων και να προληφθεί overfitting και αδυναμίες γενίκευσης/αναγωγής σε γενικευμένο πλαίσιο.

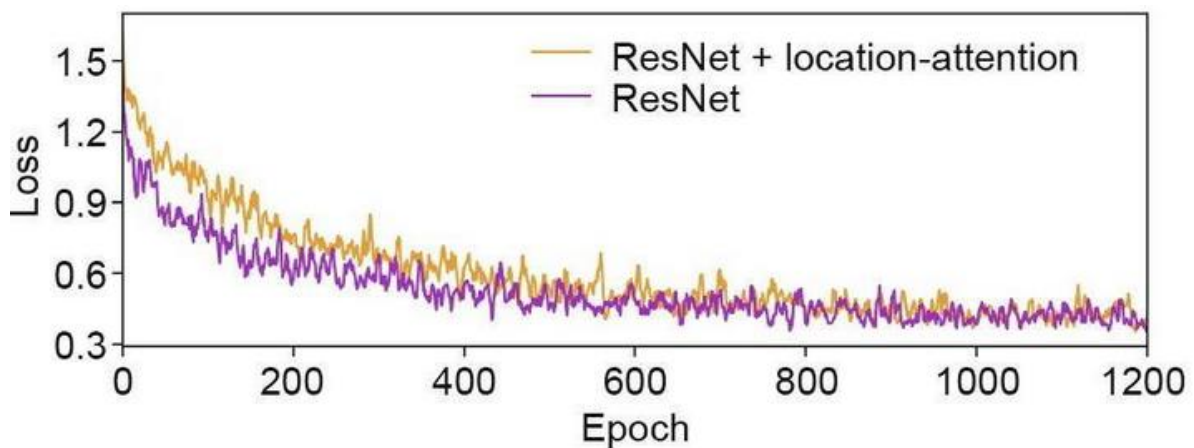
Το σύστημα στην έξοδο αποτελείται από ένα διάνυσμα 256 διαστάσεων το οποίο οδηγείται/τροφοδοτεί ένα διάνυσμα 16 διαστάσεων, μέσω δικτύου πλήρους σύνδεσης. Για το μοντέλο ταξινόμησης της θέσης (σημείο προσοχής), η τιμή της σχετικής απόστασης από τα όρια (edge), αρχικά κανονικοποιήθηκε στην ίδια τάξη μεγέθους (magnitude) και στη συνέχεια συνενώθηκε σε αυτό το πλήρους σύνδεσης δίκτυο. Τέλος υπάρχουν τρία επίπεδα πλήρους σύνδεσης για την έξοδο του τελικού αποτελέσματος με αντίστοιχη βαθμολόγηση εμπιστοσύνης.

Υλικό εφαρμογής (Hardware)

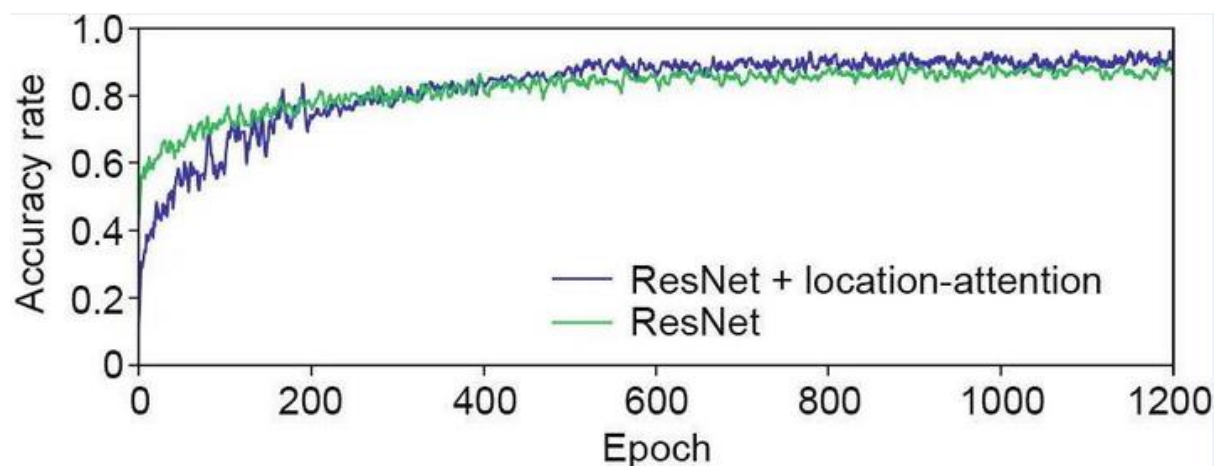
Η πλατφόρμα εφαρμογής του μοντέλου είναι επεξεργαστής i7-8700k και κάρτα γραφικών NVIDIA GeForce GTX 1080ti, ως μονάδα server. Ο χρόνος επεξεργασίας εξαρτάται σε μεγάλο βαθμό από τα επίπεδα/στρώσης της κάθε εικόνας σε ένα CT. Μέσος όρος επεξεργασίας 30 δευτερόλεπτα για ένα CT 70 επιπέδων, από την προεπεξεργασία στην έξοδο.

Διαδικασία εκπαίδευσης

Για την μέτρηση της πορείας της εκπαίδευσης του μοντέλου, καθώς και της βελτίωσης της εκπαίδευσής τους, χρησιμοποιήθηκε για μοντέλο ταξινόμησης η διασταυρούμενη εντροπία (Εικόνα 33,34).



Εικόνα 34 Η πορεία μείωσης της Loss σε Εποχές.



Εικόνα 35 Η πορεία βελτίωσης ακρίβειας σε Εποχές.

Σύμφωνα με την εικόνα 11, όταν οι εποχές (epochs) της εκπαίδευσης του προγράμματος, αυξήθηκαν άνω των 1000, η απώλεια (Loss) δεν εμφανίζει περαιτέρω μείωση ή αύξηση, το

οποίο σημαίνει ότι τα δυο μοντέλα συνέκλιναν σε μία σχετική βέλτιστη κατάσταση χωρίς διακριτή υπερπροσαρμογή (distinct overfitting). Ως μία εποχή (epochs) ορίζεται ο συνολικός αριθμός επαναλήψεων που απαιτούνται για την εκπαίδευση του μοντέλου μάθησης με όλα τα δεδομένα (εκπαίδευσης) σε ένα κύκλο. Σε μία εποχή, όλα τα δεδομένα χρησιμοποιούνται μία φορά. Δεν αναφερόμαστε σε ροή των δεδομένων, αλλά επεξεργασία αυτών και «χτίσιμο» του μοντέλου μάθησης.

Παράλληλα, στην καμπύλη ακρίβειας (εικόνα 12) της εκπαίδευσης τα δυο μοντέλα μετά από 600 εποχές (epochs) εμφανίζει το μοντέλο ResNet+location-attention μία ελαφρός υψηλότερη ακρίβεια.

Αποτελέσματα

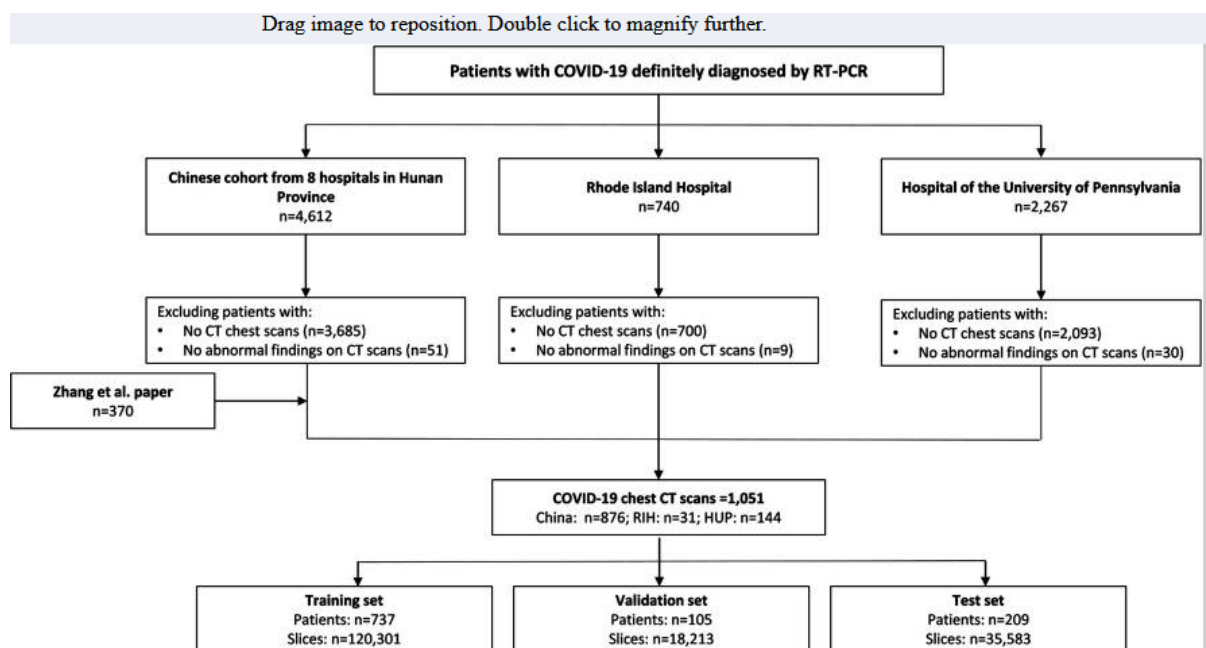
Η παρούσα μελέτη, οδήγησε στη δημιουργία μοντέλου Τεχνητής Νοημοσύνης, το οποίο εμφανίζει διακριτική ικανότητα, ικανότητα ταυτοποίησης/αναγνώρισης σε ποσοστό ακρίβειας 86,7%.

Μοντέλο ομάδας Wang et al.

Στο ίδιο προσπάθεια υπάρχει και η ομάδα του Wang et al. (Wang R., 2021), η οποία δημιούργησε ένα μοντέλο Τεχνητής Νοημοσύνης, ώστε να μελετήσουν και να είναι σε θέση για αυτοματοποιημένη διάγνωση, εικόνες από Υπολογιστική Τομογραφία (CT images).

Δεδομένα Υλικά

Τα δεδομένα που χρησιμοποιήθηκαν ήταν από 1.051 ασθενείς θετικούς σε Covid-19, διαπιστωμένο με RT-PCR. Συνολικά δεδομένα υπήρξαν από 9 νοσοκομεία των περιοχών της επαρχίας Χουνάν (Hunan) της Κίνας, το Πανεπιστημιακό Νοσοκομείο της Πενσυλβάνιας στη Φιλαδέλφεια των ΗΠΑ, το νοσοκομείο Rhode Island στο Providence, καθώς και δεδομένα που είχαν δημοσιοποιηθεί στα πλαίσια άλλων ερευνών στο ίδιο πεδίο μελέτης (Zhang K, 2020). Επίσης, στα πλαίσια διαμοιρασμού των δεδομένων, εφαρμόστηκε το νοσοκομειακό σύστημα PACS (Picture Archiving and Communications System), ενώ για της Κίνας από το Κινεζικό Εθνικό Κέντρο Βιοπληροφορικής.

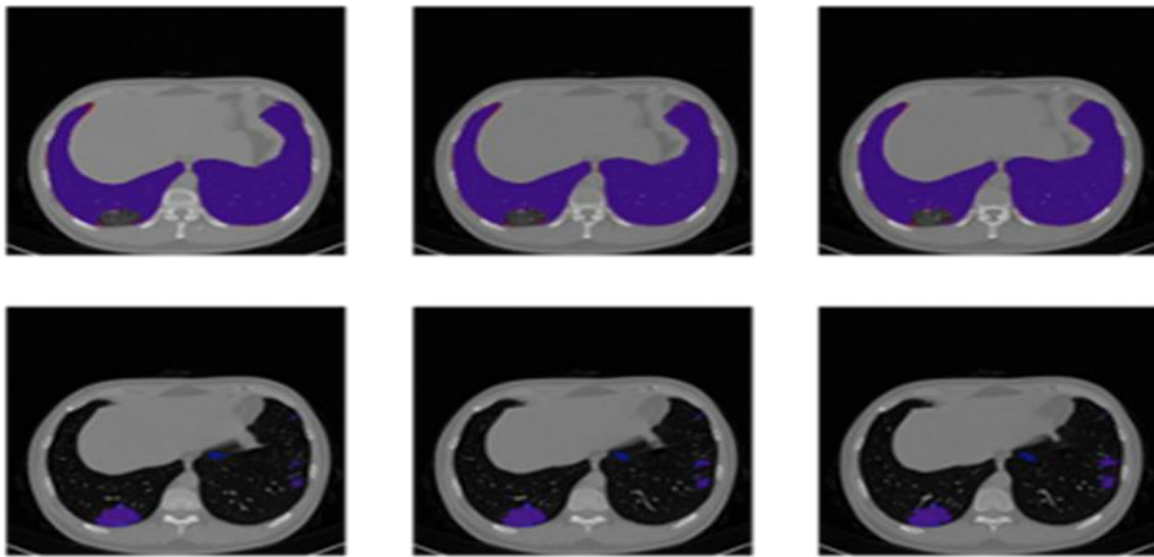


Εικόνα 36 Παρουσίαση επιλογής ασθενών.

Μοντέλο Τεχνητής Νοημοσύνης

Αρχικά, οι πνευμονικοί ιστοί καθώς του πνευμονικού παρεγχύματος, οι οποίοι είχαν αλλοιώσεις από τον Covid-19, αυτόματα ταξινομήθηκαν οι CT εικόνες με εφαρμογή ενός μοντέλου βαθιάς μάθησης (Deep Learning DL), μέσω της χρήσης βαθέως συνελεκτικού

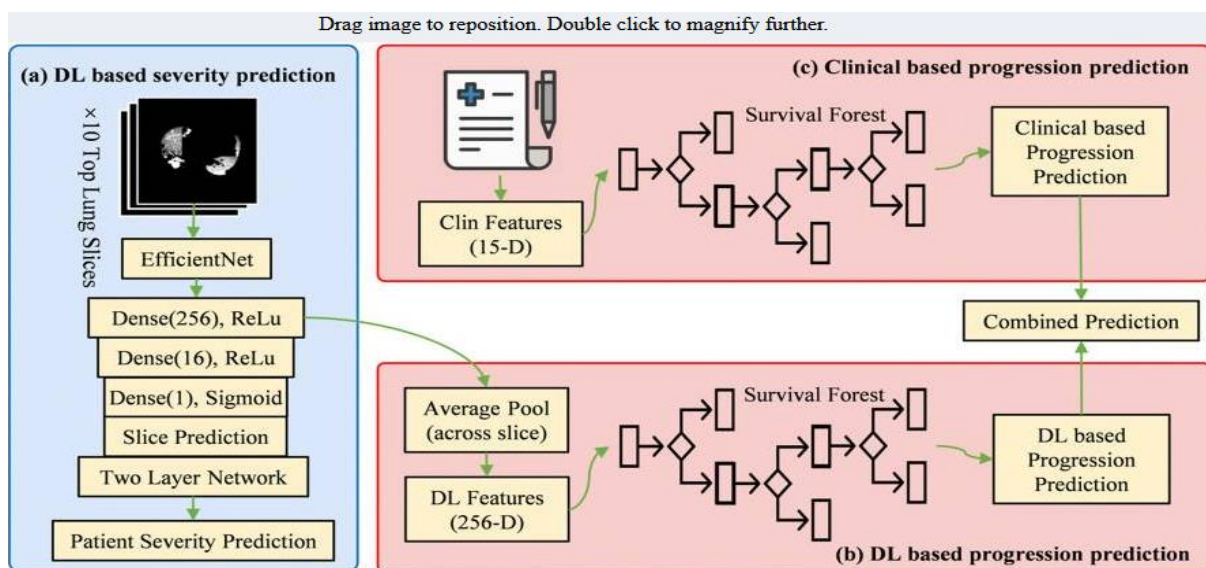
(convolutional) νευρωνικού δικτύου. Παράδειγμα της μεθόδου ταξινόμησης είναι στην εικόνα 14.



Εικόνα 37 Παρουσίαση διαχωρισμού CT εικόνων. Manual (red) and automatic (blue) lung (top row) and lesion (bottom row) segmentations on 3 slices.

Σε δεύτερο στάδιο, μοντέλο DL για τη πρόβλεψη της σοβαρότητας της εξέλιξης της πάθησης, μελετά τις εικόνες από τον πνεύμονα, ώστε να προσδιοριστεί κατά πόσο θα υπάρξει κρίσιμη ή μη αλλοίωση τους ιστούς.

Στο τρίτο στάδιο, δημιουργήθηκαν μοντέλα πρόγνωσης DL, τα οποία βασίστηκαν να έχουν είσοδο, σε δεδομένα από μοντέλα πρόγνωσης επικινδυνότητας και κλινικά δεδομένα, ώστε να δημιουργηθεί ένα βαθμολόγιο κινδύνου για τη κάθε περίπτωση. (Εικόνα 37)



Εικόνα 38 Περιγραφή του μοντέλου Τεχνητής Νοημοσύνης στην επεξεργασία και εκπαίδευση του.

Μεθοδολογία διαχωρισμού πρώτου σταδίου

Στο πρώτο στάδιο έχουμε το διαχωρισμό, ταξινόμηση των CT εικόνων, μέσω νευρωνικού δικτύου τμηματοποίησης, όπου υπάρχει η δυνατότητα εντοπισμού των βλαβών όσο και όλου του πνεύμονα. Στο βήμα αυτό έχουμε τη συμβολή τόσο ανθρώπινου παράγοντα, όπου οι πρώτες 97 περιπτώσεις τμηματοποιήθηκαν από τεχνολόγο CT και μεταπτυχιακό φοιτητή, χρησιμοποιώντας μία threshold-based και μία ημι-αυτόματη μέθοδο τμηματοποίησης, με την επίβλεψη ειδικού ραδιολόγου. Επίσης, έχουμε και τη χρήση 3D-Net, ώστε να μπορέσουν να χρησιμοποιήσουν πληροφορία από γειτονικές/ενδιάμεσες τομές.

Πραγματοποιώντας την προ-επεξεργασία, οι ακατέργαστες (raw) CT εικόνες κανονικοποιήθηκαν μεταξύ 0 και 1, ώστε να υπάρχει το ίδιο μέγεθος ογκοστοιχείου (voxel) 1mm X 1mm X 1mm. Κατά τη διάρκεια της εκπαίδευσης, επιλέχθηκε τυχαία ένα 3D πλαίσιο (patch), με μέγεθος 64 X 128 X 128, ως είσοδο για κάθε επανάληψη. Σε αυτό το βήμα έγινε χρήση διασταυρούμενες εντροπίας και της soft Dice. Κατά τη δοκιμή, το παράθυρο (sliding-window) περιοχής (patch) με μέγεθος ίδιο κατά την εκπαίδευση, εφαρμόστηκε ώστε να καλύπτει όλη τη τρισδιάστατη εικόνα. Σε αυτό το σημείο, επιτρέποντας η αλληλοεπικάλυψη περιοχών (patch), με χρήση χαρτών πιθανότητας. Στο τέλος έχουμε τις τελικές εικόνες αλλοιώσεων επαναφέρονται στην αρχική ανάλυση.

Για να αξιολογηθεί η απόδοση της τμηματοποίησης (segmentation), πραγματοποιήθηκε πείραμα, όπου οι αρχικές 97 τμηματοποιήσεις, ξαναδόθηκαν στο σύστημα τυχαία, σε ομάδες εκπαίδευσης και επικύρωσης με αναλογία 3:2. Στο τέλος υπήρξε και οπτικός έλεγχος από μεταπτυχιακούς φοιτητές καθώς και μέλη της ερευνητικής ομάδας, για να διορθωθούν τα όποια σφάλματα. Παράλληλα, υπήρξε και η τελική εποπτεία από 4 ακτινολόγους μακράς θητείας (εμπειρίας).

Μεθοδολογία μοντέλου πρόγνωσης κινδύνου

Η προεπεξεργασία των δεδομένων των εικόνων CT, έγινε με ορισμό του παραθύρου στις διαστάσεις 500-750. Το παράθυρο για το πνεύμονα μετρήθηκε σε μονάδες Hounsfield δημιουργώντας εικόνες των 8bit για κάθε 2διαστάσεων αξονικών τομών στην CT. Έπειτα, μάσκες πνεύμονα και αλλοιώσεων, που προήλθαν από τη τεχνητή νοημοσύνη, μαζί με τις επεξεργασμένες εικόνες, χρησιμοποιήθηκαν για να δημιουργήσουν την εικόνα του πνεύμονα και των αλλοιώσεων. Οι εικόνες αυτές, μετά, αναβαθμίζονται σε επίπεδο ανάλυσης 224 X 224 pixel εικόνες.

Για τη πρόβλεψη βαρύτητας, το DL μοντέλο εκπαιδεύτηκε τις 10 κορυφαίες τμηματοποιημένες τομές πνεύμονα, στη περιοχή της βλάβης, για κάθε ασθενή. Χρησιμοποιήθηκε ένας διαχωρισμός 7:1:2 σετ εκπαίδευσης (n=737) επικύρωσης (n=105) και τεστ (n=209).

Όσο αναφορά την αρχιτεκτονική έχουμε την EfficientNet B4. Στην είσοδο κάθε τμήμα (slice) στοιβάχθηκε (stacked) σε τρία κανάλια, που θα χρησιμοποιηθούν για τη προ-εκπαίδευση στα βάρη στο ImageNet. Δυο επιπλέον, πυκνά επίπεδα (256, 16) προστέθηκαν χρησιμοποιώντας ReLU (Rectified linear unit), με κανονικοποίηση και μέσο όρο επιπέδου συγκέντρωσης (pooling layer). Στο τέλος υπάρχει ένα εφαρμόζεται, μία δομή νευρώνα, όπου έχουμε ένα επίπεδο ταξινόμησης για την πρόβλεψη της σοβαρότητας για κάθε μεμονωμένη τομή (slice). Ο νευρώνας έχει αρχή σιγμοειδούς ενεργοποίησης. Οι τομές που συγκεντρώνονται από κάθε ασθενή χρησιμοποιούνται ως είσοδο για ένα νευρωνικό δίκτυο 2 επιπέδων πλήρους σύνδεσης, για την πρόβλεψη της σοβαρότητας σε κάθε ασθενή. Τέλος χρησιμοποιήθηκε ένας στοχαστικός βελτιστοποιητής μείωσης της κλίσης με ρυθμό εκμάθησης 0.0001 και μέγεθος ομάδας (batch) 16.

Μέθοδος Πρόβλεψης Πρόγνωσης

Η ικανότητα για την πρόβλεψη της πρόγνωσης της πάθησης καθώς και του κινδύνου αυτού είναι κρίσιμος παράγοντας και δυνατότητα για ένα μοντέλο Τεχνητής Νοημοσύνης. Στη συγκεκριμένη εφαρμογή έχουμε χρήση δεδομένων τόσο κλινικών όσο και δεδομένων που προήλθαν από DL, για τη πρόβλεψη της σοβαρότητας της πάθησης. Η μέθοδος έχει δύο κλάδους.

Για την πρόβλεψη πρόγνωσης βασισμένη σε DL, χρησιμοποιήθηκε μία σειρά από 256-διαστάσεων DL, από τις 10 κορυφαίες CT τομές επιλέγονται βασισμένες στη μεγαλύτερη περιοχή αλλοίωσης, σε αξονική προβολή, όπου εξάγεται από το πυκνό (256) επίπεδο των προ-εκπαίδευσης μοντέλων πρόγνωσης κινδύνου.

Στη συνέχεια, αυτά τα δυανύσματα των χαρακτηριστικών, τα οποία συγκεντρώθηκαν ως μέσο όρο σχηματίζοντας τα χαρακτηριστικά του DL, το οποίο θα αποτελέσει είσοδο για ένα «δένδρο» επιβίωσης και άρα μία βαθμολόγηση κινδύνου για κάθε ασθενή μέσα στα δεδομένα, των οποίων το μοντέλο, προβλέπουν το κίνδυνο εξέλιξης.

Για τη κλινική πρόβλεψη εξέλιξης, θα χρησιμοποιηθούν ιατρικές και βιολογικές μεταβλητές όπως η ηλικία, το φύλο, ο πυρετός, ο αριθμός των λευκών αιμοσφαιρίων, ο αριθμός των λεμφοκυττάρων, ύπαρξη καρδιαγγειακής νόσου, η ύπαρξη υπέρτασης, η ύπαρξη Χρόνιας Αποφρακτικής Πνευμονοπάθειας, ο διαβήτης, η εμφάνιση ηπατικής νόσου, νεφροπάθειας, καρκίνος, HIV, και κυρίως επαφή με φορέα COVID-19. Αυτά θα εξαχθούν μέσω δημογραφικών καταγραφών. Αυτά τα κλινικά δεδομένα τροφοδοτούν ένα έτερο μοντέλο δένδρο επιβίωσης/κινδύνου ώστε να εξαχθούν οι τελικές βαθμολογίες κινδύνου. Έπειτα, για κάθε ασθενή, ο βαθμός κινδύνου από το DL και από τα κλινικά δεδομένα, προστίθενται δίνοντας μαζί σε συνδυασμό το κίνδυνο εξέλιξης της ασθένειας.

Στατιστική ανάλυση

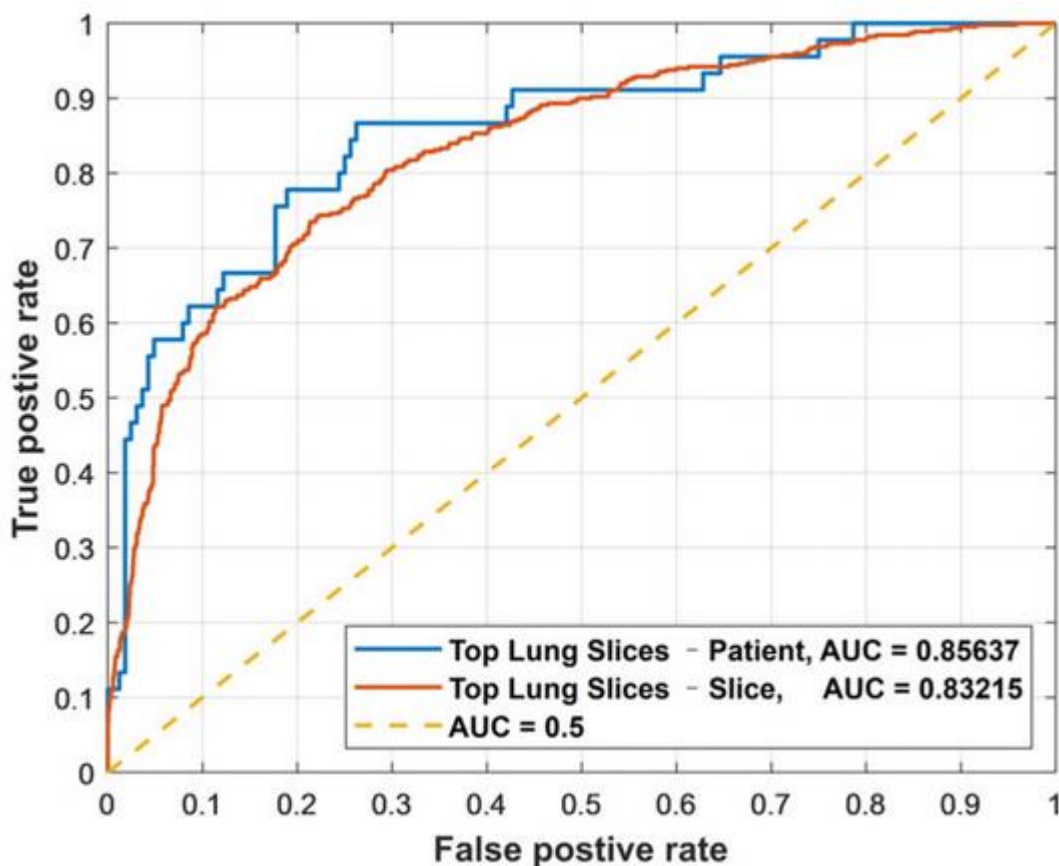
Η αξιολόγηση των τμηματοποιήσεων έγινε με υπολογισμό τον συντελεστή ομοιότητας Dice καθώς και οπτικής εξέτασης. Η ακρίβεια, η ευαισθησία, η ειδικότητα και η περιοχή κάτω από την χαρακτηριστική καμπύλη λειτουργίας του δέκτη (ROC-AUC, receiver operating characteristic curve), υπολογίστηκε για τα μοντέλα ταξινόμησης. Στο 95% των αξιόπιστων δεδομένων στην ακρίβεια, την ευαισθησία και την ειδικότητα, εφαρμόζεται η μέθοδος Wald (Agresti A, 1998). Ο C-index στα ψευδώς θετικά, εφαρμόστηκε για την αξιολόγηση της απόδοσης των μοντέλων πρόβλεψης (Harrell FE, 1996). Τέλος το χρονικά εξαρτώμενο ROC-AUC υπολογίστηκε από τις βαθμολογίες κινδύνου και τις πληροφορίες εξέλιξης μέσω της μεθόδου Kaplan-Meier (Heagerty PJ, 2005), ώστε να υπάρξει επιπλέον αξιολόγηση της απόδοσης πρόβλεψης της εξέλιξης.

Υλικό εφαρμογής (Hardware)

Τα μοντέλα Βαθιάς Μάθησης (DL) εφαρμόστηκαν σε λειτουργικό Keras (version 2.2.5) και Tensorflow (version 1.12.3) και εκπαιδεύτηκαν σε NVIDIA V100 GPUs. Τα μοντέλα πρόβλεψης προόδου εφαρμόστηκαν σε Scikit-Learn (version 0.21.3). Ο κώδικας είναι διαθέσιμος στη σελίδα <http://github.com/robinwang08/COVID19>.

Απεικόνιση απόδοσης μοντέλου

Κατά την εφαρμογή του μοντέλου, στα δείγματα ασθενών, έγινε χρήση των 10 καλύτερων δειγμάτων πνεύμονα, με τη μεγαλύτερη περιοχή βλάβης. Το σύστημα για κάθε ασθενή πέτυχε πολύ υψηλά ποσοστά ακρίβειας στο επίπεδο της τομής (0.778 – 95% CI: 0.76-0.795), ευαισθησίας (0.71 - 95% CI: 0.667–0.751) και ειδικότητας (0.796 - 95% CI: 0.776–0.815). Παράλληλα σε συγκεντρωτικό επίπεδο το μοντέλο έχει ακρίβεια (0.833 - 95% CI: 0.776–0.877), ευαισθησία (0.622 - 95% CI: 0.476–0.749) και ειδικότητα (0.89 - 95% CI: 0.832–0.930). Σύνοψη των άνω αποτελεσμάτων είναι στο γράφημα ROC-AUC (Εικόνα 33).



Εικόνα 39 Απεικόνιση απόδοσης μοντέλου.

Αποτελέσματα

Το μοντέλο κατάφερε στη πρόγνωση κινδύνου να αγγίξει το ποσοστό 95% επαλήθευσης με τα ιατρικά/κλινικά δεδομένα, μία πολύ μεγάλη εγκυρότητα. Καθώς επίσης 97% στον εντοπισμό των πάσχων περιοχών στις ct εικόνες (Wang R., 2021).

Εφαρμογές στην ανακάλυψη φαρμάκων

De novo σχεδιασμός χημικών μορίων ενάντια στον SARS-CoV-2

Πρόκειται για την εργασία της ομάδας των Bung et al. (Bung N., 2021), η οποία δημοσιεύτηκε το 2021 και με τη χρήση τεχνητής νοημοσύνης, κατάφερε να δημιουργήσει έναν αλγόριθμο για τον προσδιορισμό φαρμακευτικών μορίων ενάντια της πρωτεάσης 3CL του ιού, της οποίας η λειτουργία είναι ζωτικής σημασίας για τον ιό και μπορεί έτσι να αδρανοποιηθεί.

Συλλογή δεδομένων

Τα δεδομένα για την εκπαίδευση του βαθέως νευρωνικού μοντέλου συλλέχθηκαν από τη βάση δεδομένων ChEMBL (Gaulton A, 2012). Μία ομάδα δεδομένων της τάξης του 1.6 εκατομμυρίων φαρμακευτικών μικρών μορίων συλλέχθηκαν. Επειδή υπήρχε περιορισμένη γνώση για η καταστολή της πρωτεάσης 3CL από μικρά μόρια, δεδομένα από μόρια, τα οποία πειραματικά έχουν αποδειχθεί ότι καταστέλλουν ικές πρωτεάσες επιλέχθηκαν. Αυτό οδήγησε σε 7665 μόρια.

Από αυτά τα μόρια, μέσω του pChEMBL, και με τη χρήση του AutoDock Vina (Trott O, 2010), βρέθηκαν ποια από αυτά είχαν βαθμολογία άνω του 7.0, για το ενεργό κέντρο της 3CL πρωτεάσης του SARS-CoV-2. Στο τέλος, κατέληξαν 2515 μόρια τα οποία θα τροφοδοτήσουν το νευρωνικό δίκτυο. Πρέπει να σημειωθεί ότι όλες οι βάσεις δεδομένων μικρών μορίων εφαρμόζουν το Simplified Molecular Input Line Entry System (SMILES) (Weininger D, 1989), για να αξιοποιηθεί η αποτελεσματικότητα των επαναλαμβανόμενων νευρωνικών δικτύων στο χειρισμό σειράς (sequential) δεδομένων.

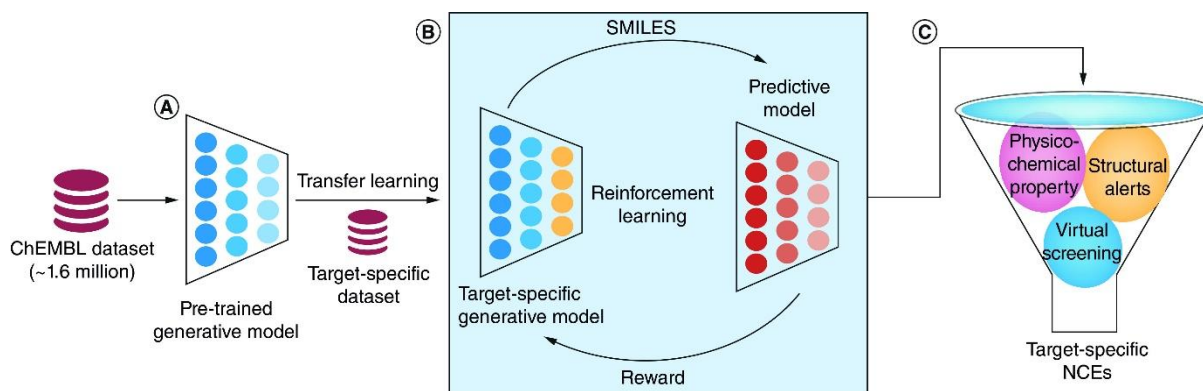
Προεπεξεργασία δεδομένων

Οι ομάδες δεδομένων του SMILES προεπεξεργάστηκαν, εφαρμόζοντας φίλτρα, ώστε να απομακρυνθούν μόρια τα οποία στερεοχημικά, ή λόγω αλατότητας ή λόγω ατόμων μη επιθυμητών. Με τη διαδικασία αυτή, αφαιρέθηκαν SMILES strings με περισσότερα από 100 στοιχεία σε μήκος, με αποτέλεσμα το 97% των ομάδων δεδομένων που έμειναν ήσαν μικρότερες από των <100 συμβόλων. Στο τέλος οι ομάδες κανονικοποιήθηκαν για την αφαίρεση περιττών μορίων. Στη διαδικασία χρησιμοποιήθηκε η βιβλιοθήκη RDKit της Python. Επίσης όλα τα SMILES strings στις ομάδες δεδομένων (dataset) σημειώθηκαν με έναν

χαρακτήρα έναρξης αλληλουχίας και έναν χαρακτήρα λήξης αλληλουχίας και για την κωδικοποίηση (one-hot encoded) των SMILES strings χρησιμοποιήθηκαν 39 σύμβολα.

Μαθαίνοντας την γλώσσα των μικρών μορίων με τη χρήση παραγόμενων μοντέλων

Ομάδα μορίων σε πλήθος περίπου 1.6 εκατομμυρίων μικρών μορίων σε SMILES μορφή (format) χρησιμοποιήθηκαν για την προ-εκπαίδευση του γενετικού μοντέλου (Εικόνα 34). Η αρχιτεκτονική του βαθύς νευρωνικού δικτύου του παραγόμενου μοντέλου αποτελείται από ένα μονό (single) επίπεδο των 1024 επαναλαμβανόμενων ομάδων αμφίδρομων πυλών (Gated Recurrent Units GRUs) (Chung J, 2014), ως εσωτερική μνήμη, και επιπλέον μία στοίβα, η οποία λειτουργεί ως εξωτερική μνήμη (Joulin A, 2015). Η επιπρόσθετη στοίβα των υπάρχοντων GRUs κελιών βελτιώνει την ικανότητα την επαναλληψιμότητα του νευρωνικού δικτύου, στην καταγραφή των συντακτικών και σημασιολογικών χαρακτηριστικών εγγενώς, στο context-free grammar των διαδοχικών δεδομένων (sequential data). Η εκπαίδευση πραγματοποιήθηκε εφαρμόζοντας mini-batch μειούμενης/καθοδικής κλίσης με AMSGrad βελτιστοποιητή (optimizer) (PT, 2019).



Εικόνα 40 De novo βήματα σχεδιασμού φαρμάκων, για τη δημιουργία μικρών μορίων ενάντια σε στόχο.

Κατά τη φάση των συμπερασμάτων, ο χαρακτήρας της έναρξης της ακολουθίας, δόθηκε ως είσοδο το γενετικό μοντέλο και οι επόμενοι χαρακτήρες του SMILES sting χρησιμοποιήθηκαν ένας κάθε φορά εφαρμόζοντας πολυωνυμική δειγματοληψία. Η διαδικασία δειγματοληψίας είτε εάν τερματίζεται είτε με την εμφάνιση του χαρακτήρα τέλους της ακολουθίας είτε εάν το μήκος της συμβολοσειράς SMILES υπερέβαινε ένα προκαθορισμένο μέγιστο όριο. Η χημική ευστάθεια ελέγχθηκε με την βιβλιοθήκη RDKit για κάθε δείγμα της SMILES string για να

εξασφαλιστεί η συνθετική εφαρμοσιμότητα (feasibility), του μικρού μορίου που δημιουργήθηκε. Το μοντέλο εκπαιδεύτηκε για 500 εποχές (epochs) σε κάρτα γραφικών Tesla V100 (GPU) και τα βάρη από το εκπαιδευόμενο μοντέλο χρησιμοποιήθηκαν για τις επόμενες στη σειρά διεργασίες (Εικόνα 33). Όλες οι υλοποιήσεις έγιναν στο PyTorch (Paszke A, 2019).

Σχεδιασμός φαρμάκων ligand-based χρησιμοποιώντας μεταφορά μάθησης

Το προεκπαιδευτικό γενετικό μοντέλο εκπαιδεύτηκε για τη δημιουργία νέων μικρών μορίων που μοιάζουν με φάρμακα χωρίς στοχοειδική πληροφορία. Στη συνέχεια, οι πληροφορίες του στόχου, ενσωματώθηκαν στο μοντέλο μέσω της ομάδας δεδομένων για τα στοχοειδικά μικρά μόρια. Το προεκπαιδευτικό γενετικό μοντέλο με γνώση της γραμματικής των SMILES επανεκπαιδεύτηκε, ειδικά σε αυτά τα μικρά μόρια που μπορούν να συνδεθούν με τη πρωτεΐνη στόχο (Εικόνα 34B).

Στη μεταφορά μάθησης (transfer learning TL), η κατανομή των πιθανοτήτων που μαθαίνει το σύστημα, από τα τελευταία επίπεδα αλλάζει προκειμένου να οδηγήσει το γενετικό μοντέλο, εστιασμένα σε ένα μικρότερο υποσύνολο του χημικού χώρου, μέσα στον οποίο θα γίνουν οι όποιες χημικές αλληλεπιδράσεις. Το μοντέλο εκπαιδεύτηκε με το ίδιο σύνολο υπερπαραμέτρων για 100 εποχές (epochs) σε Tesla K20 GPU. Η κατανομή για το μέγιστο συντελεστή Tanimoto (AH, 1999), για τα δημιουργημένα μικρά μόρια με τα εκπαιδευόμενα δεδομένα (training dataset) επισημαίνεται (plotted) κάθε 10 εποχές (epochs) για την παρακολούθηση της μεταφοράς μάθησης (TL). (Εικόνα 33).

Βελτιστοποίηση ιδιοτήτων με βάση το μοντέλο πρόβλεψης

Ένα σύνολο δεδομένων (dataset) μικρών μορίων με ειδικότητα για το στόχο και οι αντίστοιχες τιμές των ιδιοτήτων τους μπορούν να χρησιμοποιηθούν για εκπαίδευση στο προγνωστικό μοντέλο. Τα SMILES strings υποβάλλονται σε προεπεξεργασία μέσω των ίδιων φίλτρων που χρησιμοποιήθηκαν στο γενετικό μοντέλο. Το βαθύ νευρωνικό δίκτυο του προγνωστικού μοντέλου (Εικόνα 34), αποτελείται από αμφίδρομες GRUs ως επιπλέον μοντέλο μνήμης. Επίσης χρησιμοποιήθηκε ένα επίπεδο ενσωμάτωσης για τη μετατροπή των SMILES strings, στην είσοδο, σε ένα πυκνότερη μορφή πριν την εισαγωγή τους στο GRU επίπεδο. Δυο πυκνά επίπεδα γραμμικής διόρθωσης (linear rectified) χρησιμοποιούνται μετά τα αμφίδρομα GRU επίπεδα και την εξαγωγή των τιμών των ιδιοτήτων. Κάθε αμφίδρομο GRU επίπεδο

εναλλάσσεται με ένα επίπεδο απόρριψης (dropout layer) για να αποφευχθεί η υπερπροσαρμογή (overfitting) καθώς το μοντέλο είναι σε φάση εκπαίδευσης.

Οι υπερπαραμέτροι ρυθμίζονται προσεκτικά για επιτευχθεί το ελάχιστο λάθος στη πρόγνωση. Το μοντέλο εκπαιδεύτηκε χρησιμοποιώντας μικρο-ομάδες (mini-batch) μειούμενης βαθμίδας (κλίσης) με βελτιστοποιητή Adam (Kingma DP, 2014). Το μοντέλο έτρεξε 500 εποχές (epochs) σε Tesla V100 GPU.

Συνδυασμός του παραγόμενου και προγνωστικού μοντέλου με τη χρήση ενισχυτικής μάθησης. Το παραγόμενο μοντέλο, δηλαδή μοντέλο TN που βασίζεται σε στατιστική πρόβλεψη του επόμενου στοιχείου μιας ακολουθίας ή δομής, που προέκυψε μετά τη μεταφορά μάθησης (TL) βελτιστοποιήθηκε περαιτέρω χρησιμοποιώντας ενισχυτική μάθηση (Reinforcement Learning, RL). Χρησιμοποιήθηκε μία κανονικοποιημένη μέθοδος στην εκπαίδευση του παραγόμενου μοντέλου χωρίς καταστροφική παράβλεψη (forgetting) των χαρακτηριστικών που αποκτήθηκαν στη προεκπαίδευση και στην μεταφορά μάθησης (TL).

Κατά τη διάρκεια του RL, το στοχο-ειδικό γενετικό μοντέλο και το προγνωστικό μοντέλο δρουν ως πράκτορας και κριτής (agent and critic), αντίστοιχα. (Εικόνα 33B). Ο πράκτορας έχει ήδη μάθει κατά την προεκπαίδευση μία κύρια πολιτική, η οποία είναι η κατανομή των πιθανοτήτων στα διάφορα σύμβολα σε κάθε θέση της συμβολοσειράς SMILES. Η διαδικασία πρόβλεψης του επόμενου συμβόλου, δεδομένων των προηγούμενων, και της κρυφής κατάστασης του μοντέλου, πραγματοποιείται από τον πράκτορα. Κάθε ενέργεια του πράκτορα συμβάλλει σε μία τμηματική (episodic) ανταμοιβή (reward), η οποία υπολογίζεται στο τέλος της μεμονωμένης δειγματοληψίας. Ο στόχος της μεθόδου είναι να βελτιώσει ή να βελτιστοποιήσει την υπάρχουσα πολιτική, με γνώμονα η ανταμοιβή (reward) να είναι η μέγιστη.

Η ανταμοιβή (reward) είναι μία τιμή που υπολογίζεται χρησιμοποιώντας μια συνάρτηση ανταμοιβής και ορίζεται από την προβλεπόμενη ιδιότητα της συμβολοσειράς SMILES, όπως αυτή υπολογίζεται μέσω του προγνωστικού μοντέλου. Εάν η τιμή είναι εντός επιθυμητού εύρους, έχουμε ανταμοιβή, διαφορετικά ποινή.

Η εν λόγω εκπαίδευση χρησιμοποιήθηκε για την εξαγωγή των μορίων (Εικόνα 33), όπου ως βελτιστοποιητή εφαρμόστηκε ο AMSGrad (PT, 2019). Το μοντέλο έτρεξε 50 εποχές (epochs) σε Tesla K20 GPU.

Φιλτράρισμα των παραγόμενων μορίων μέσω φυσικοχημικών ιδιοτήτων

Τα παραγόμενα μικρά μόρια που κανονικοποιήθηκαν και αφαιρέθησαν τα όποια διπλότυπα καθώς και μόρια όμοια με το ChEMBL σύνολο μορίων εκπαίδευσης. Στη συνέχεια υποβλήθηκαν σε επιπλέον φιλτράρισμα, με γνώμονα φυσικοχημικές ιδιότητες και τις ιδιότητες της περιοχής του μορίου στόχου. Σε αυτό το σημείο ήταν χρήσιμη η χρήση γνωστών αναστολέων για τον εν λόγω στόχο και οι φυσικοχημικές τους ιδιότητες.

Τα διάφορα φίλτρα που εφαρμόστηκαν περιελάμβαναν μία σύνθετη βαθμολογία : προσβασιμότητας (για τη περιοχή) <5.0 , ποσοτική εκτίμηση φαρμακευτικής ομοιότητας (drug-likeness) >0.4 , χρήση του συντελεστή οκτανόλης-νερού (logP) <6.0 , βαθμολογία βιοδραστικότητας >6.0 και μοριακό βάρος 400-800Da. Τα μόρια που δημιουργήθηκαν πέρασαν όλα τα φίλτρα τα επόμενα στάδια. (Εικόνα 33).

Εφαρμογή φίλτρων για αφαίρεση μορίων με ανεπιθύμητες ομάδες

Στην ομάδα των μορίων που πέρασαν και το φιλτράρισμα από τις φυσικοχημικές ιδιότητες, εφαρμόστηκε επιπλέον επιλογή με τη χρήση των φίλτρων Pan Assay Interference Compounds (Baell JB, 2010), BRENK (Brenk R, 2008), NIH (Doveston RG, 2015) και ZINC. Αυτά τα φίλτρα εφαρμόζονται για την αποφυγή τοξικών υποομάδων στα μόρια καθώς και τεχνητά ανέφικτων υποομάδων στα μόρια. Το RDKit χρησιμοποιήθηκε για την εφαρμογή και των 4 φίλτρων στα μόρια, και όποιο μόριο σημαίνονταν τουλάχιστον από δυο φίλτρα αφαιρούνταν. Οι υπόλοιπες ενώσεις ακολούθησαν στο επόμενο στάδιο μοριακής μοντελοποίησης.

Μοριακή μοντελοποίηση

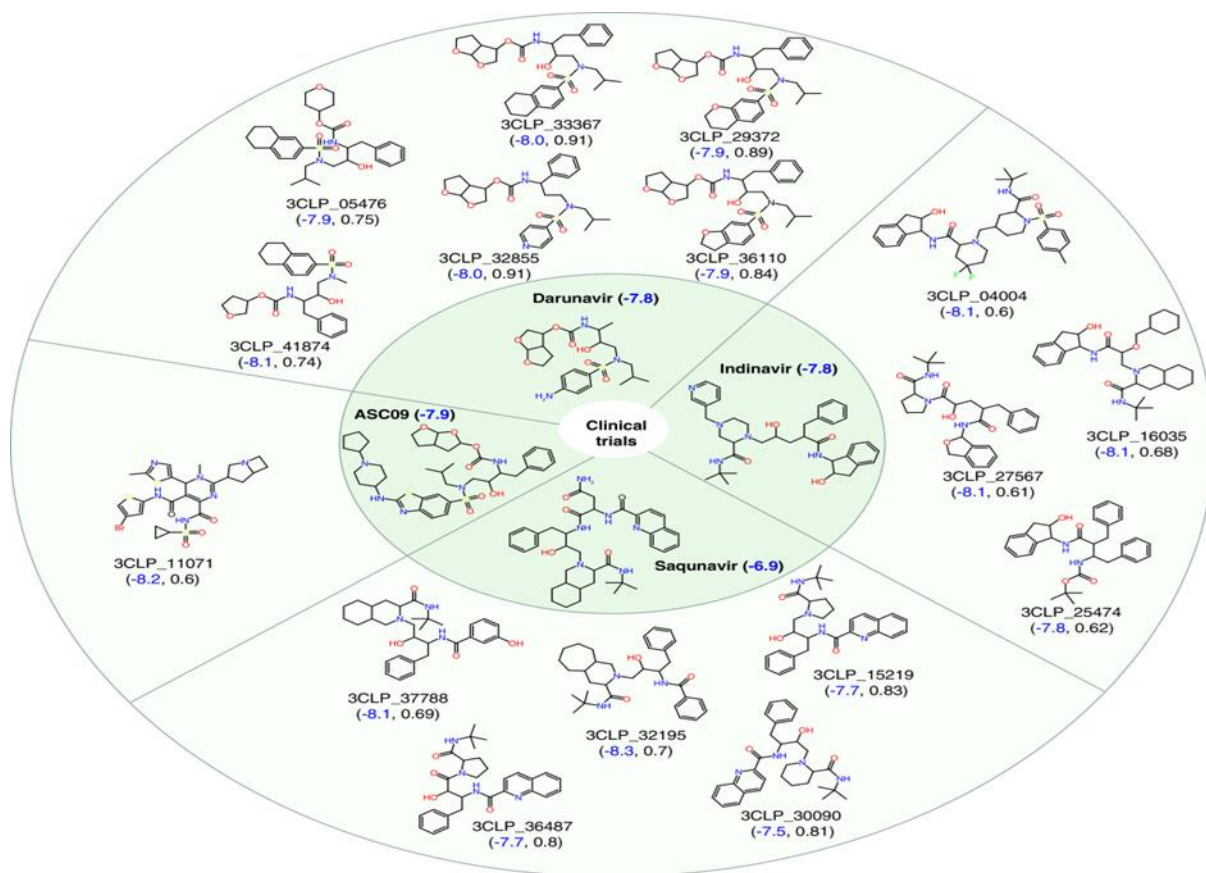
Το τελικό στάδιο είναι η μοριακή μοντελοποίηση χρησιμοποιώντας το AutoDock Vina (Trott O, 2010). Εφαρμόστηκε βαθμολογία για τη συγγένεια δέσμευσης των μορίων με τη πρωτεΐνη στόχο. Τα μόρια με βαθμολογία <7.0 θεωρήθηκαν ως πιθανά NCEs ενάντια στη πρωτεάση του SARS-CoV-2. (Εικόνα 333).

Αποτελέσματα

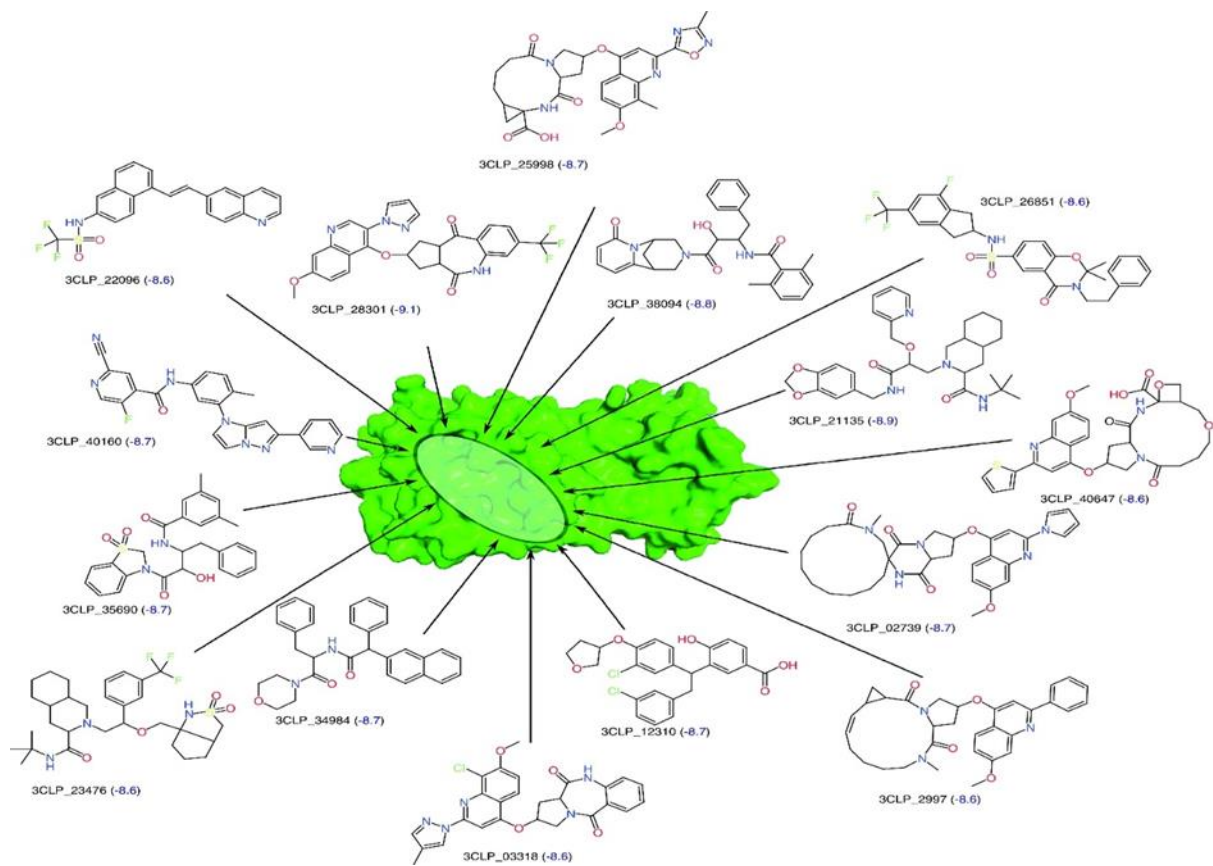
Στόχος αυτού του συστήματος, ήταν η εύρεση κατάλληλων καταστολέων πρωτεάσης του Covid-19. Μετά το στάδιο της Ενισχυμένης Μάθησης (RL), επιλέχθηκαν αρχικά 50.000 μόρια. Από αυτά, αφού αφαιρέθηκαν τα διπλότυπα και τα ChEMBL-identical, έμειναν 42.484 μόρια. Οι φυσικοχημικές ιδιότητες αυτών μορίων υπολογίστηκαν με τη χρήση της βιβλιοθήκης RDKit. Τα παραγόμενα μόρια είχαν παρόμοιες ιδιότητες με τις πρωτεάσες ειδικά μόρια, κατά το στάδιο της TL.

Στη συνέχεια με την εφαρμογή των φυσικοχημικών φίλτρων, μόνο 3960 μόρια κατάφεραν να περάσουν και τα 5 φυσικοχημικά φίλτρα/ιδιότητες.

Τέλος το AutoDock Vina, το οποίο εφαρμόστηκε για τον HIV καταστολέα της πρωτεάσης (C, 2020), χρησιμοποιήθηκε για virtual screening. Ως αποτέλεσμα μας έδωσε φάρμακα που έμοιαζαν και θα μπορούσαν να έχουν υψηλή ομοιότητα με τους καταστολείς πρωτεάσης του HIV. (Εικόνα 36). Ενώ στη συνέχεια έχουμε τη τελική επιλογή μορίων με τη μέγιστη βαθμολογία (Εικόνα 37).



Εικόνα 41 Βαθμολόγηση μορίων και ομαδοποίηση σε σχέση με καταστολείς πρωτεάσης του HIV.



Εικόνα 42 Τελική επιλογή και εφαρμογή των μορίων/καταστολέων, σε σχέση με την δράση τους στη πρωτεάση του Covid-19.

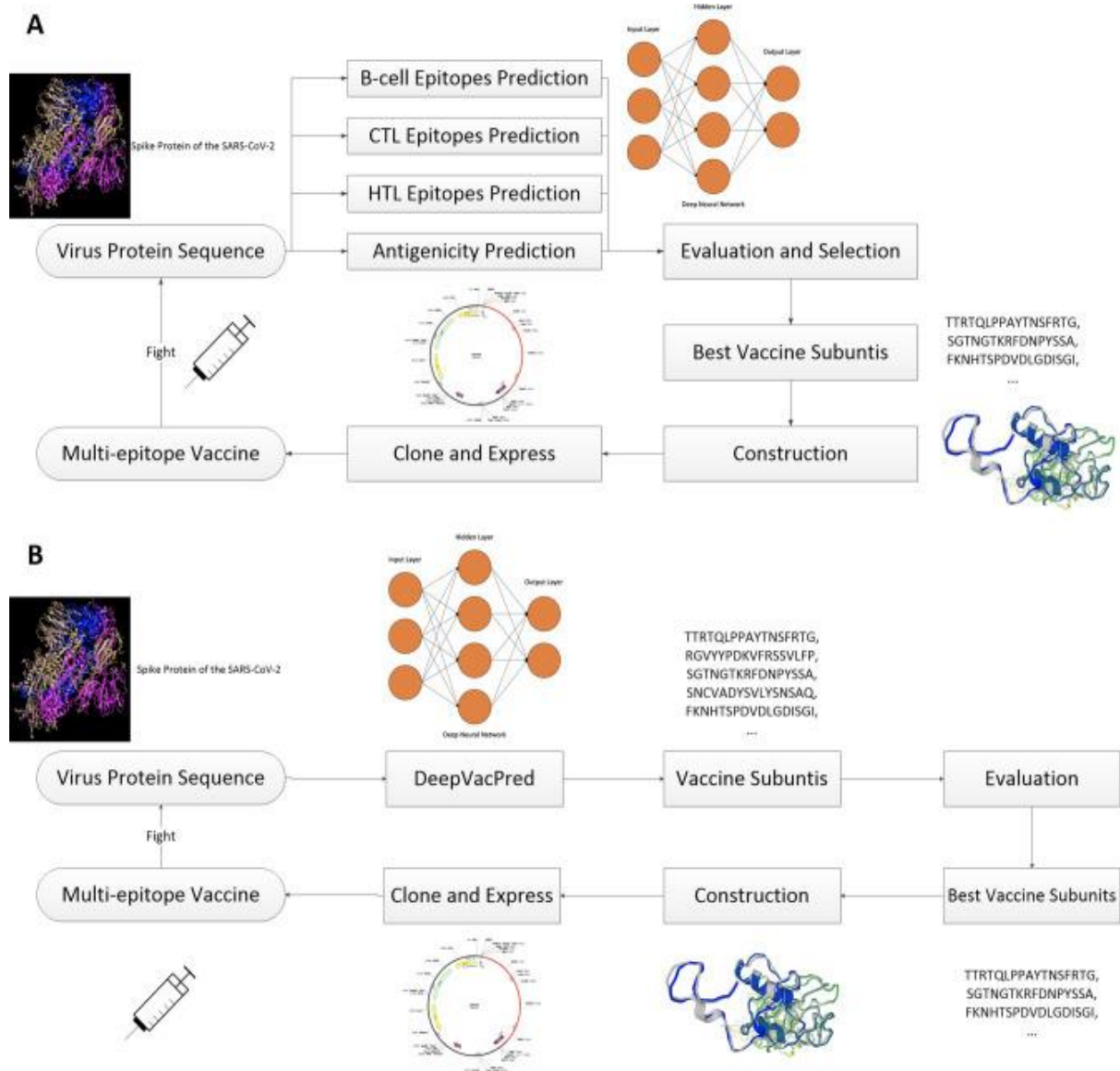
Εφαρμογή στην ανακάλυψη εμβολίων

Ένα ιδιαίτερο πεδίο χρήσης της τεχνικής νοημοσύνης είναι και η ανάγκη για ταχεία ανάπτυξη εμβολίου ενάντια στην επιδημία του SARS-CoV-2. Σε αυτή τη προσπάθεια άξια προσέγγιση και άξια μελέτης είναι η δημιουργία *in silico* βαθιάς μάθησης για την εύρεση πολλαπλών επιτόπων για τα εμβόλια, από την ομάδα Yang Z, et al. (Yang Z., 2021).

DeepVacPred

Η δημιουργία των εμβολίων βασίζεται στο να μετατραπεί ο ιός και οι πρωτεΐνες της επιφάνειάς του (επίτοποι), σε μεμονωμένα μόρια (πρωτεΐνες) ή δομές αυτών αναγνωριζόμενα από τα αντισώματα του οργανισμού, αλλά χωρίς να υπάρχει δυνατότητα τοξικής ή λοιπής παθογόνου δράσης στον οργανισμό. Όστε σε περίπτωση επαναμόλυνσης από τον παθογόνο οργανισμό ή ιό να μπορεί ο οργανισμός, λόγω κυτταρικής μνήμης, να τον αντιμετωπίσει πολύ γρηγορότερα, δίχως εμφάνιση συμπτωμάτων.

Στη προσπάθεια αυτή υπήρξε ο σχεδιασμός του μοντέλου βαθύς νευρωνικού δικτύου το οποίο μας δίνει τη δυνατότητα να μπορέσουμε να σχεδιάσουμε/επιλέξουμε επιτόπους που μπορούν να χρησιμοποιηθούν για τη κατασκευή εμβολίων. Με την εφαρμογή του DeepVacPred βελτιώνεται το ένα ήδη υπάρχον μοντέλο χρήσης υπολογιστικής βιολογίας/νευρωνικού δικτύου (Εικόνα 42).



Εικόνα 43 Σχεδιάγραμμα της Υπολογιστικής Σχεδίασης εμβολίων/επιτόπων.

Στην εικόνα 36A, έχουμε το υπάρχον μοντέλο (χωρίς το DeepVecPred) όπου βλέπουμε ότι έχουμε τη «παραδοσιακή» διαδικασία υπολογιστικού σχεδιασμού εμβολίου. Ειδικότερα, γίνεται χρήση των εργαλείων για να μπορέσει να γίνει πρόβλεψη επιτόπων στα Β-κύτταρα, τα CTL και HTL για όλες τις πρωτεΐνες του ιού (Purcell A, 2007) (E., 2020). Η αντιγονικότητα καθώς και άλλες φυσικοχημικές ιδιότητες των επικαλυπτόμενων τμημάτων του ιού είναι επίσης χρήσιμες και λαμβάνονται υπόψιν και αξιολογούνται. Κάθε εργαλείο μπορεί να προβλέψει μόνο ένα στόχο/επίτομο στόχο τη φορά και όχι σε όλες τις σειρές κυττάρων. Για παράδειγμα το εργαλείο BerIPred (Jespersen M, 2017) είναι για επιτόπους στα Β κύτταρα και στα Τ κύτταρα εφαρμόζεται το εργαλείο NetMHCpan (Nielsen M, 2007).

Απέναντι σε αυτή την αδυναμία, αναπτύχθηκε το DeepVecPred, το οποίο, μέσω βαθέως νευρωνικού δικτύου, μπορεί να μελετήσει την πρωτεΐνη ακίδα του SARS-CoV-2, να επιλέξει 26 πιθανούς στόχους πάνω στην εν λόγω πρωτεΐνη και στη συνέχεια να προβλέψει τους επιτόπους των Β κυττάρων, τους επιτόπους CTL και τους HTL, ώστε να κατασκευαστεί ο εμβόλιο. Παράγοντες όπως η πληθυσμιακή κάλυψη, οι φυσικοχημικές ιδιότητες, η αντιγονικότητα, η αλλεργιογένεση και η τοξικότητα λαμβάνονται υπόψη/αξιολογούνται, καθώς επίσης η δευτεροταγής και η τριτοταγής δομή. Τέλος υπάρχει και η υπολογιστική μελέτη του γονιδιώματος/κωδικονίων για βελτίωση της έκφρασής του, όπως και τα σενάρια μεταλλάξεων (Yang Z., 2021).

Δεδομένα / υλικά

Για τη δημιουργία και εκπαίδευση του μοντέλου, χρησιμοποιήθηκαν 5000 επίτοποι Β-κυττάρων (B) (οι πλέον πρόσφατοι) και 2000 επίτοποι Τ-κυττάρων (T) που περιέχονται στο Μείζον Σύμπλεγμα Ιστοσυμβατότητας (ΜΣΙ – Major Histocompatibility Complex – MHC) -1 και -2 (Ahmad T, 2016), από την IEDB (Immune Epitope Database – Βάση Ανοσολογικών Επιτόπων), καθώς επίσης σε ίδιους αριθμούς πρωτεΐνες που δεν είναι επίτοποι των Τ και Β-κυττάρων, δημιουργώντας μία «βάση δεδομένων» επιτόπων και μη. Παράλληλα, 100 πρόσφατα γνωστά προστατευτικά αντιγόνα (από την IEDB) και ίδιος αριθμός πρωτεϊνών με μη-προστασία τυχαίας επιλογής, σε συνδυασμό με 400 αντιγόνα από άλλη έρευνα (Heinson A, 2017). Στο σύνολο 600 αντιγόνα (Yang Z., 2021).

Οργάνωση Δεδομένων

Το DeepVecPred δημιουργήθηκε ως δομή υποβοηθούμενης (supervised) μάθησης πάνω σε σωστά οργανωμένο (subtly) όγκο δεδομένων. Για να μπορέσει το μοντέλο, να εξάγει υποψήφια μόρια, θα πρέπει η πρωτεϊνική αλληλουχία θα πρέπει να περιέχει τουλάχιστον ένα επίτοπο από τα Τ-κύτταρα και έναν από τα Β-κύτταρα και παράλληλα να είναι προστατευτικά αντιγόνα. Με αυτό το τρόπο δημιουργούνται ομάδες (set) Cartesian Product (O., 1995), τα οποία περιέχουν όλους τους συνδυασμούς άνω συνδυασμούς σε δυο ομάδες. Με αυτό το τρόπο τα TxB, και BxT που δημιουργούνται έχουν όλους τους δυνατούς συνδυασμούς των γνωστών επιτόπων Τ και Β-κυττάρων.

Σε αυτό εφαρμόζεται η χρήση/εκπαίδευση με τα 600 αντιγόνα του νευρωνικού δικτύου που μπορεί να αναγνωρίζει προστατευτικά αντιγόνα. Συνολικά εμφανίζονται 706.970 αλληλουχίες πεπτιδίων και είναι το σύνολο αναφερόμενα σε Β-κύτταρα και Τ-κύτταρα (Yang Z., 2021).

Εκπαίδευση Δικτύου

Το νευρωνικό δίκτυο του DeepVacPred είναι ένα πολλαπλών επιπέδων περιελεγμένο (convolutional) νευρωνικό δίκτυο (CNN – convolutional neural network) και ένα 4-επιπέδων γραμμικό νευρωνικό δίκτυο (linear neural network) συνδεδεμένα, δημιουργώντας ένα βαθύ νευρωνικό δίκτυο με 2 ταξινομήσεις (class) στην έξοδο. Εμφανίζονται δυο ομάδες (set) θετικά και αρνητικά, οι οποίες επισημούνται (annotated) με τη μέθοδο Z-descriptors (Hellberg S, 1987), και μετατρέπονται σε ίσου μήκους 45 διανύσματα (vectors) με μεταστροφή τύπου ACC (auto cross covariance) (Wold S, 1993). Το δίκτυο εκπαιδεύεται από τα νέα δεδομένα (της μεταστροφής), το DNN (Deep Neural Network) επιτυγχάνει να ταξινομήσει τη λειτουργία και να προβλέψει εάν στην είσοδο είναι ένα προστατευτικό αντιγόνο το οποίο περιέχει επιτόπους και από τα Β-κύτταρα και τα Τ-κύτταρα, και να αποφασίσει εάν η συγκεκριμένη αλληλουχία μπορεί να αποτελέσει υποψήφιο μόριο για το εμβόλιο.

Αυτό το DNN είναι το βασικό μέρος της ταχείας διαδικασίας σχεδιασμού εμβολίου και ονομάζεται DNN-V. Παράλληλα, υπάρχει και το δίκτυο για τους επιτόπου των Τ-κυττάρων με όνομα DNN-T.

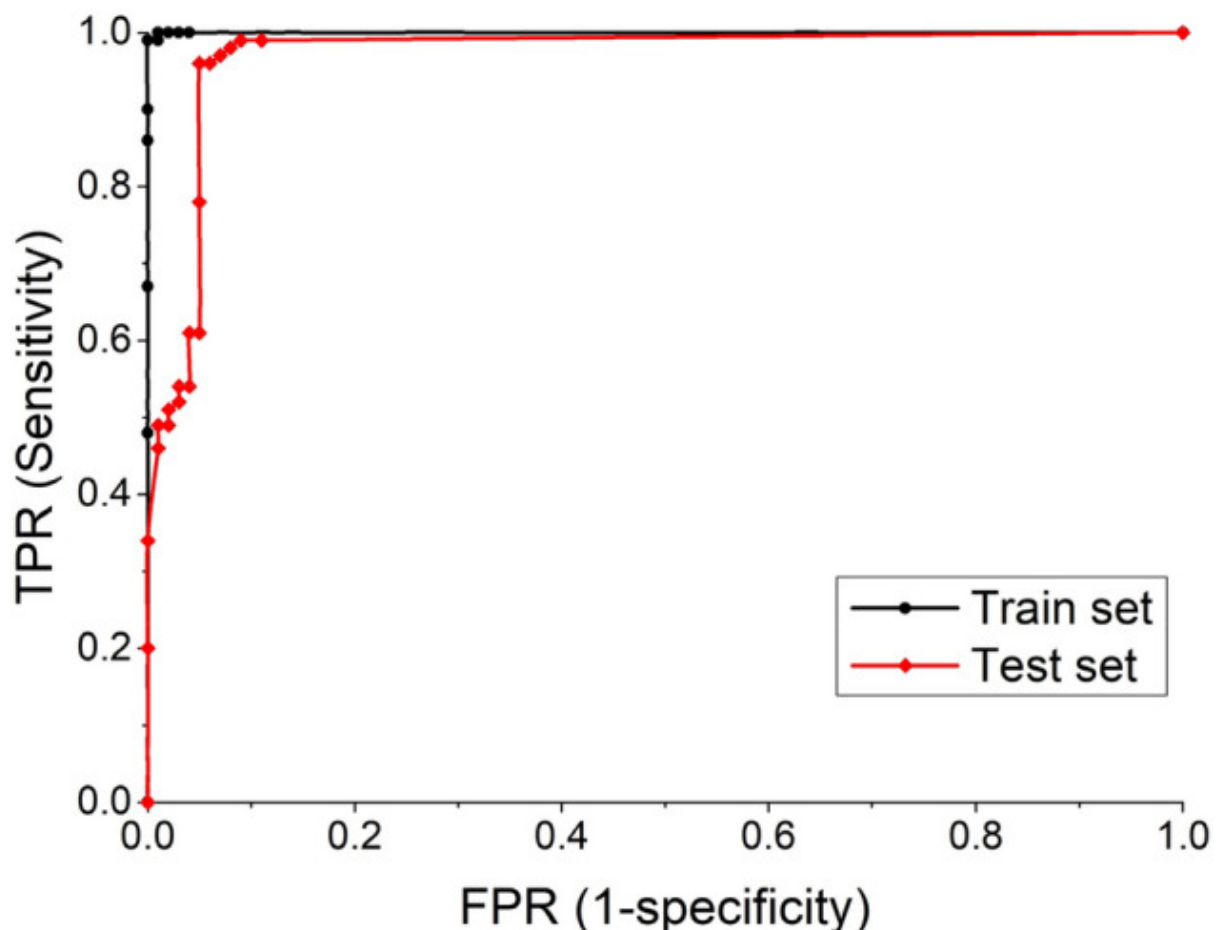
Επαλήθευση

Για την επαλήθευση της λειτουργίας του μοντέλου, χρησιμοποιείται η καμπύλη ROC (Receiver Operating Characteristic) (Calders T, 2007), η οποία θα απεικονίσει κατά πόσο το σύστημα μπορεί, όντως να εφαρμόσει τη διαγνωστική του λειτουργία του DNN-V (Εικόνα 37).

Το σύστημα θα δοκιμαστεί σε δυο σύνολα δεδομένων, ένα της εκπαίδευσης και ένα της εξέτασης (test), όπου το καθένα περιέχει 200 αλληλουχίες πρωτεϊνών. Η ομάδα πρωτεϊνών εκπαίδευσης περιέχει 200 τυχαία επιλεγμένες από το σύνολο των δεδομένων και στην εκπαίδευση του DNN-V περιέχει 100 θετικές και 100 αρνητικές αλληλουχίες πρωτεϊνών.

Επίσης, έγινε επιλογή γνωστών επιτόπων Β-κυττάρων και Τ-κυττάρων, οι οποίοι δεν εμπεριέχονται στα επιλεγμένα δεδομένα, όπου χρησιμοποιήθηκαν στα άνω βήματα εκπαίδευσης, επίσης με 100 θετικές και 100 αρνητικές αλληλουχίες πρωτεϊνών.

Τα δεδομένα επικύρωσης εμφανίζονται στην εικόνα. Τα όρια είναι από 0 έως 1. Η ακρίβεια που αναφέρεται στον πίνακα είναι η μεγαλύτερη τιμή όλων των ορίων. Οι τιμές ευαισθησίας και ειδικότητας στον Πίνακα είναι για την περίπτωση της υψηλότερης ακρίβειας. Η τιμή AUC (Area Under the ROC Curve) είναι 0.9703 και το σενάριο των δοκιμών, το οποίο υποδεικνύει την υψηλή ακρίβεια της ταξινόμησης του DNN-V για τον εντοπισμό πιθανών υπομονάδων εμβολίου.



Εικόνα 44 Καμπύλη επαλήθευσης μοντέλου.

Η περιοχή κάτω από τις καμπύλες αντιπροσωπεύει την ικανότητα του DNN-V να ταξινομεί πιθανές υπομονάδες εμβολίου και μη πιθανές. Η περιοχή κάτω από τις καμπύλες ROC υποδηλώνει ότι το DNN-V έχει ισχυρή ικανότητα ταξινόμησης και υψηλή ακρίβεια στις περισσότερες τιμές του κατωφλίου.

Πίνακας 1 Επαλήθευση του DeepVecPred.

Είδος διαδικασίας	AUC	Κατώφλι (Threshold)	Ακρίβεια (Accuracy)	Ευαισθησία Sensitivity	Ευαισθησία Specificity
Εκπαίδευση	0.9999	0.32	0.995	0.99	0.99
Δοκιμή	0.9703	0.5	0.95	0.95	0.95

Η υψηλότερη ακρίβεια είναι στη τιμή 0.995 (στην ομάδα εκπαίδευσης) και η τιμή κατωφλίου (Threshold) 0.32. Σε τέτοια τιμή κατωφλίου η ειδικότητα είναι 0.99. Για το σύνολο των δοκιμών η υψηλότερη ακρίβεια είναι 0.95, τιμή κατωφλιού 0.5, και ευαισθησία και η ειδικότητα με τιμή 0.95. Τα πειράματα δείχνουν υψηλή ακρίβεια και ισχυρή ικανότητα ταξινόμησης του DeepVecPred.

Ανάλυση Μοντέλου

Σχεδίαση και εκπαίδευση του DNN στο DeepVecPred framework

Κάθε στοιχείο εισόδου στο DNN είναι μία αλληλουχία με μήκος 45 διανυσμάτων, και σε κάθε μία πρωτεΐνη εφαρμόζεται η z-description (Hellberg S, 1987) καθώς και ο μετασχηματισμός ACC (Wold S, 1993). Το Συνελικτικό Νευρωνικό Δίκτυο (CNN) εμφανίζει καλή απόδοση στον εντοπισμό και επεξεργασία των διανυσμάτων, ενώ το πολλαπλών επιπέδων γραμμικό νευρωνικό δίκτυο, το οποίο είναι συνδεδεμένο με την έξοδο του CNN, σχηματίζει ένα σύνθετο DNN και ενισχύει τη ταξινομική ικανότητα του μοντέλου. Ως αποτέλεσμα, το DNN με τα πολλαπλά επίπεδα και τις παραμέτρους του κάθε επιπέδου, αποφασίζει εφαρμόζοντας τυχαία αναζήτηση ώστε να αυτήν με τη υψηλότερη ακρίβεια αλλά και αποδεκτή ταχύτητα εκτέλεσης λύση:

Οι ορισμένες υπερπαραμέτροι εκπαίδευσης τους DNN είναι οι κάτωθι:

- i. Learning rate: [0.0001, 0.0005, **0.001**, 0.0015, 0.002];
- ii. Optimizer: [SGD, RMSProp, **Adam**];
- iii. Epochs: [2000, 4000, **6000**, 8000, 10000];
- iv. Batch size: [1024, 2048, **4096**, 8192].

Γραμμική πρόβλεψη επιτόπων B-κυττάρων

Για αυτό το βήμα έγινε χρήση (Yang Z., 2021):

- 1) BepiPred-2.0 web server ([BepiPred-2.0](#)) : εργαλείο μηχανικής μάθησης το οποίο βασίζεται σε αλγόριθμο random forest και το set εκπαίδευσης καλύπτει ένα μεγάλο αριθμό από γνωστούς γραμμικούς B-κυττάρων επιτόπους, προέλευσης IEDB (Jespersen M, 2017).
- 2) ABCpred ([ABCpred](#)): εφαρμόζει επαλαμβανόμενο νευρωνικό δίκτυο για τη ταξινόμηση των επιτόπων καθώς και των μη-επιτόπων, ώστε να βελτιώσει την ακρίβεια (Saha S, 2006).
- 3) SVMTrip ([SVMTrip](#)): χρησιμοποιεί μηχανή διανυσμάτων υποστήριξης για να προβλέψει τους αντιγονικούς επιτόπους και θέτει AUC είναι στη τιμή 0.702 (Yao B, 2012).
- 4) BCPreds ([BCPreds](#)):είναι μηχανή διανυσμάτων υποστήριξης με AUC τιμή 0.758 και η πρόβλεψή του βασίζεται σε μεθόδους πυρήνα (El-Manzalawy Y, 2008). Η προσβασιμότητα της επιφάνειας των B λεμφοκυττάρων ελέγχεται με το λογισμικό IEDB Emini (Emini EA, 1985).

Πρόβλεψη επιτόπων T κυτταροτοξικών λεμφοκυττάρων

Έγινε χρήση του NetMHCpan 4.1 server ([NetMHCpan](#)) για τη πρόβλεψη των CTL επιτόπων, για κάθε πιθανό επίτοπο υποψήφιο για εμβόλιο. Η πρόβλεψη έγινε με επιτόπους CTL με μήκος 9aa. Όλες οι παράμετροι προκαθορισμένες (default). Το NetMHCpan προβλέπει τη δέσμευση του πεπτιδίου σε οποιοδήποτε μόριο MHC-I, γνωστής αλληλουχίας, με την εφαρμογή τεχνητού νευρωνικού δικτύου, το οποίο εκπαιδεύεται με περισσότερα από 850000 από Binding Affinity (BA) και Mass-Spectrometry Eluted Ligands (EL) πεπτίδια (Yang Z., 2021) (Jurtz V, 2017).

Πρόβλεψη επιτόπων T βοηθητικών λεμφοκυττάρων

Έγινε χρήση του NetMHCIIpan4.0 server ([NetMHCIIpan](#)). Η πρόβλεψη έγινε για επιτόπους T βοηθητικών κυττάρων μήκους 15aa. Όλες οι παράμετροι προκαθορισμένες (default). Το NetMHCIIpan προβλέπει τη σύνδεση για όλα τα MHCII, με τη χρήση νευρωνικού δικτύου, το οποίο εκπαιδεύεται με 500000 επιτόπους BA, EL, HLA-DR, HLA-DQ και HLA-DP (Reynisson B, 2020).

Πρόβλεψη δευτεροταγούς δομής

Έγινε χρήση του PSIPRED ([psipred](#)), το οποίο ενσωματώνει δυο νευρωνικά δίκτυα τα οποία δέχονται ως είσοδο τα δεδομένα από το PSI-BLAST (Position Specific Iterated-BLAST). Επιτυγχάνεται μέσος όρος βαθμολογίας Q3 81.6% (McGuffin LJ, 2000). Επίσης γίνεται χρήση του RaptorX Property ([predict](#)) για την πρόβλεψη της προσβασιμότητας (ACC) και των περιοχών (DISO). Το Raptor X χρησιμοποιεί το μοντέλο μηχανικής μάθησης DeepCNF (Deep Convolutional Neural Fields) για τη πρόβλεψη της δευτεροταγής δομής, του ACC και DISO (Wang S L. W., 2016).

Εμβολίου τρισδιάστατο (3D) μοντέλο δομής

Για το 3D μοντέλο στο τελικό εμβόλιο έγινε χρήση του RaptorX server ([ContactMap](#)). Το RaptorX μπορεί να δώσει την αναδίπλωση των πρωτεϊνών μέσω ενός μοντέλου αποστάσεων σε βάση βαθειάς μάθησης. Αυτός ο server, στις CASP12, CASP13, κατετάγη 1^{ος} στην απόδοση για τη πρόβλεψη δομής πρωτεϊνών από βαθιά μάθηση (Wang S L. W., 2016).

Επιβεβαίωση της 3D δομής του εμβολίου

Το τελικό τρισδιάστατο μοντέλο του εμβολίου επικυρώνεται με τη χρήση του ProSA-web server ([prosa.php](#)), το οποίο μπορεί να υπολογίσει τη συνολική βαθμολογία της ποιότητας για μία συγκεκριμένη δομή που δέχεται ως είσοδο. Εάν αυτή η βαθμολογία είναι πέρα από ένα εύρος τιμής για τις εγγενείς πρωτεΐνες, η δομή πιθανόν να έχει σφάλματα. Μια γραφική παράσταση τοπικών βαθμολογιών υποδεικνύει τα σημεία του μοντέλου που έχουν πρόβλημα, και μία τρισδιάστατη απεικόνιση διευκολύνει την ανίχνευση (Wiederstein M, 2007).

Αποτελέσματα

Στόχος είναι ο εντοπισμός επιτόπων που θα αποτελέσουν αντικείμενο για τα αντισώματα ενάντια στον ιό και ειδικότερα της ακίδας. Το DeepVacPred κατέληξε σε πολλαπλούς επιτόπους μήκους 649aa κατά του SARS-CoV-2. Περιέχει 11 υπομονάδες με 16 επιτόπους B-κυττάρων, 82 επιτόπους CTL και 89 επιτόπους HTL. Δείχνει καλή αντιγονικότητα, μεγάλη πληθυσμιακή κάλυψη και αποδεκτές φυσικοχημικές ιδιότητες και δομές,

Συμπεράσματα Συζήτηση

Ο covid-19 αποτέλεσε τη πρώτη μεταπολεμική επιδημία, με παγκόσμιο αντίκτυπο. Η διασπορά του, το άγνωστο του παθολογοανατομικού υποβάθρου, το άγνωστο της σωστής και έγκαιρης θεραπείας ουσιαστικά σόκαρε την ανθρωπότητα. Αποτέλεσμα ήταν ο παγκόσμιος περιορισμός ελεύθερων μετακινήσεων, υπηρεσιών και γενικότερα των άμεσων επαφών των ανθρώπων.

Παράλληλα, τα κράτη, ο Παγκόσμιος Οργανισμός Υγείας και οι μεγάλες φαρμακοβιομηχανίες ενώθηκαν, ώστε για πρώτη φορά να υπάρξει μαζικά συντονισμένη χρηματοδότηση της μελέτης του ιού, μάθησης της φύσης του, της δράσης του, της παθολογίας που προκαλεί καθώς και την ανάπτυξη κατάλληλων φαρμάκων και κυρίως εμβολίων. Ουσιαστικά, από τον Δεύτερο Παγκόσμιο Πόλεμο έχει να δει η ανθρωπότητα μία τέτοια κινητοποίηση.

Σε αυτή τη προσπάθεια χρησιμοποιήθηκε και η Τεχνητή Νοημοσύνη, ως βοήθεια στους υγειονομικούς επιστήμονες (ιατρούς και βιολόγους). Αποτέλεσμα έχουμε τις άνω μελέτες που παρατέθηκαν.

Ειδικότερα, η μελέτη Xu et al. (Xu X., 2020) , είναι εφαρμογή τεχνητής νοημοσύνης για το σχεδιασμό ενός δικτύου ταξινόμησης για τη διάκριση του Covid-19 από το IAVP. Όσο αφορά τη δομή του δικτύου είναι κλασική εφαρμογή ResNet, Σε αυτό έγινε σύγκριση μεταξύ μοντέλων με και χωρίς επιπλέον μηχανισμό εντοπισμού θέσης. Το πείραμα έδειξε ότι ο αναφερόμενος μηχανισμός, μπορεί να διακρίνει καλύτερα τις περιπτώσεις Covid-19 από άλλες. Επιπλέον, πολλαπλές μέθοδοι βελτίωσης εφαρμόστηκαν όπως η image patch vote και η Noisy-OR Bayesian function, προκειμένου να προσδιοριστούν οι κυρίαρχοι τύποι μόλυνσης. Αυτές οι προσθήκες, βελτίωσαν σταθερά τη μέση βαθμολογία F1 και το ποσοστό ακρίβειας.

Ωστόσο, αυτή η μελέτη έχει περιορισμούς. Πρώτον, οι εκδηλώσεις του Covid-19 μπορεί να έχουν κάποια αλληλοεπικάλυψη με τις εκδηλώσεις άλλων πνευμονιών, όπως η IAVP, η οργανική πνευμονία και η ηωσινοφιλική πνευμονία. Εδώ έγινε σύγκριση μόνο με την IAVP. Η κλινική διάγνωση Covid-19 πρέπει ακόμα να συνδυάζει το ιστορικό επαφής του ασθενούς, το ιστορικό των ταξιδιών, τα πρώτα συμπτώματα και τις εργαστηριακές εξετάσεις. Δεύτερον, ο αριθμός των δειγμάτων εκπαίδευσης και δοκιμής θα πρέπει να επεκταθεί για να βελτιωθεί η ακρίβεια στο μέλλον. Θα πρέπει να διεξαχθούν περισσότερες πολυκεντρικές κλινικές μελέτες για την αντιμετώπιση της σύνθετη κλινικής κατάστασης. Επιπλέον, πρέπει να καταβληθούν προσπάθειες για τη βελτίωση του μοντέλου κατάτμησης και ταξινόμησης. Ένα καλύτερο αποκλειστικό μοντέλο θα μπορούσε να σχεδιαστεί για εκπαίδευση, η ακρίβεια τμηματοποίησης και ταξινόμησης του μοντέλου θα μπορούσε να βελτιωθεί και η απόδοση

γενίκευσης του αλγορίθμου θα μπορούσε να επαληθευτεί με ένα μεγαλύτερο σύνολο δεδομένων.

Η ομάδα των Wang et al. (Wang R., 2021), δημιούργησε ένα σύστημα τεχνητής νοημοσύνης το οποίο μελετά οξονικές τομογραφίες θώρακος και δίνει τη δυνατότητα στους ιατρούς να διαχειριστούν τους ασθενείς, καθώς ενισχύει τα κλινικά δεδομένα, όσο αφορά στο να προβλέψει το κίνδυνο της εξέλιξης της κατάστασης των ασθενών. Στην μελέτη αυτή δόθηκε προσοχή στο ακριβές χρονικό διάστημα μεταξύ της εκτέλεσης της αξονικής τομογραφίας θώρακος και της πρώιμης εμφάνισης κρίσιμων γεγονότων (όπως εισαγωγή στη ΜΕΘ, διασωλήνωση ή θάνατος). Αυτό έγινε με την εφαρμογή ενός μιας πολυδιεθνούς ομάδας ασθενών από διάφορα ιδρύματα. Το σύστημα της τεχνητής νοημοσύνης κατασκευάστηκε σε ένα πλαίσιο ανάλυσης της χρονικής στιγμής (επιβίωσης) που μπορεί να χειριστεί αποτελεσματικά δεδομένα, για τη πρόβλεψη του κινδύνου. Με την εφαρμογή αυτού του μοντέλου, μπορεί να γίνει πρόβλεψη κινδύνου βασισμένη στις CT τομογραφίες και κλινικών δεδομένων και να γίνει μία κατηγοριοποίηση των ασθενών σε διάφορες ομάδες κινδύνου εξέλιξης. Επίσης μπορεί να οριστεί ένα κρίσιμο χρονικό περιθώριο για έγκαιρη θεραπεία και να υπάρξει χρονοδιάγραμμα περί της χρήσης εξοπλισμού αναπνευστικής υποστήριξης.

Ωστόσο, αυτή η μελέτη έχει και περιορισμούς. Ειδικότερα, υπάρχει η υποψία πιθανή μεροληψία της επιλογής των ασθενών, λόγω της αναδρομικής και πολυϊδρυματικής φύσης της μελέτης. Για παράδειγμα, τα κλινικά χαρακτηριστικά μεταξύ των ομάδων ασθενών στις ΗΠΑ και στην Κίνα διέφεραν. Επίσης το ποσοστό των κρίσιμων ασθενών ήταν στο 22% για την Κίνα, στις ΗΠΑ για τις δυο ομάδες ήταν στο 50%. Η ετερογένεια των δεδομένων πιθανόν αντανακλά τις διαφορές στο μοτίβο εφαρμογής (practice). Στον κινεζικό πληθυσμό, οι αξονικές τομογραφίες χρησιμοποιήθηκαν συχνότερα για τον προσυμπτωματικό Covid-19, ειδικά για τη πρώιμη περίοδο της πανδημίας. Άλλες εξηγήσεις περιλαμβάνουν τη διαφορά σε δημογραφικά στοιχεία και τον επιπολασμό της νόσου. Ωστόσο, η μελέτη αυτή περιλαμβάνει δεδομένα από 2 νοσοκομεία των ΗΠΑ, 9 νοσοκομεία της Κίνας καθώς και από άλλες μελέτες. Καταλαβαίνουμε, ότι θα πρέπει να υπάρχει μεγαλύτερος αριθμός ασθενών, καθώς και να ληφθούν σε αυτούς, υπόψιν κλινικές καταστάσεις όπως η αναπνευστική ανεπάρκεια, το σηπτικό σοκ και ο πολυοργανική ανεπάρκεια. Η παρούσα μελέτη επικεντρώθηκε σε ασθενείς που πήγαν στη ΜΕΘ, όπου διασωληνώθηκαν ή πέθαναν.

Με την έξαρση της πανδημίας, έγινε άμεσα απαιτητή η εύρεση φαρμάκων για τον Covid-19. Σε αυτή τη δράση κινήθηκε και η ομάδα του Bung et al. (Bung N., 2021), όπου μέσα από την

ανάπτυξη του συστήματος νευρωνικού δικτύου. Με τη χρήση TL και RL και την εφαρμογή αυστηρών φυσικοχημικών φίλτρων, κατάφερε το σύστημα να εξάγει μόρια τα οποία μπορούν να επιτελέσουν το ρόλο αυτό. Στη τελική επιλογή έγινε και η χρήση/κλήτευση υπαρχουσών μορίων καταστολέων πρωτεασών HIV, ώστε να υπάρξει η τελική επιλογή.

Η χρησιμότητα ενός τέτοιου λογισμικού έγκειται στο χρόνο και στην ειδικότητα. Η παραδοσιακή μέθοδος επιλογής διαρκεί πολλά χρόνια και απαιτεί άμεση, σταδιακή και συνεχή εφαρμογή σε ζωντανά κύτταρα. Αντίθετα με τον υπολογιστικό τρόπο έχουμε γρήγορη χρήση της υπάρχουσας γνώσης μορίων, και εύρεση νέων που να έχουν τις επιθυμητές ιδιότητες. Όσο αναφορά τα τελικά μόρια που επιλέχθηκαν, αυτά συγκρίθηκαν με υπάρχουσα μόρια του SymMap βάσης δεδομένων, όπου εμφανίστηκε υψηλή συγγένεια 2 μορίων με υπάρχουσα μόρια (Aurantiamide), η οποία προέρχεται από το βότανο *Baphicacanthus cusia*, και χρησιμοποιείται στο κρυολόγημα, στο πυρετό και στην γρίπη. Το φυτό υπάρχει στην Κίνα, στο Μπαγκλαντές, την Ινδία και τη Μιαμάρ. Στην Ινδία εξάγεται και από το *Piper aurantiacum*.

Στο πλαίσιο της δημιουργίας εμβολίων, η ομάδα Yang Z. et al. (Yang Z., 2021), έναντι του Covid-19, κινήθηκε στην δημιουργία ενός νέου μοντέλου νευρωνικού δικτύου, το οποίο θα εντοπίζει επιτόπους έναντι των αντιγονικών περιοχών της πρωτεΐνης ακίδας του ιού. Μέχρι τότε, η τεχνητή νοημοσύνη, εφαρμόζονταν, αλλά παράλληλα από ανεξάρτητα εργαλεία υπολογιστικής βιολογίας. Σε αντίθεση με αυτά, οι Yang Z. et al. ενοποίησαν μία μέθοδο υπολογισμού λαμβάνοντας υπόψιν στο ίδιο σύστημα και επιτόπους στα T λεμφοκύτταρα, και στα B λεμφοκύτταρα και στα T κυτταροτοξικά λεμφοκύτταρα καθώς και προστατευτικών αντιγόνων. Παράλληλα, έχουμε συνδυασμό εργαλείων που μελετούν τη δευτεροταγή αλλά και τη τριτοταγή δομή των επιτόπων, όπως και την χωροταξική αλληλεπίδραση των πρωτεϊνών (docking) και των μοριακών αλληλεπιδράσεων μεταξύ τους.

Αυτή η κίνηση και σχεδιασμός ήταν καινοτόμος. Μέσω αυτού του εργαλείου μπορούμε να έχουμε μείωση κατά 95% των περιττών προβλέψεων αντιγόνων που μπορούν να χρησιμοποιηθούν/μελετηθούν για το εμβόλιο ενάντια στον Covid-19. Επίσης με πολύ μεγάλη ταχύτητα, διότι το DeepVacPred μπορεί να υπολογίσει αυτές τις 26 υπομονάδες του εμβολίου πολύ μικρό χρονικό διάστημα, σχεδόν σε δευτερόλεπτα. Επομένως ένας ερευνητής μπορεί να κατασκευάσει ένα εμβόλιο πολλαπλών επιτόπων για έναν νέο ιό και να επικυρώσει την ποιότητά του σε διάστημα μίας ώρα.

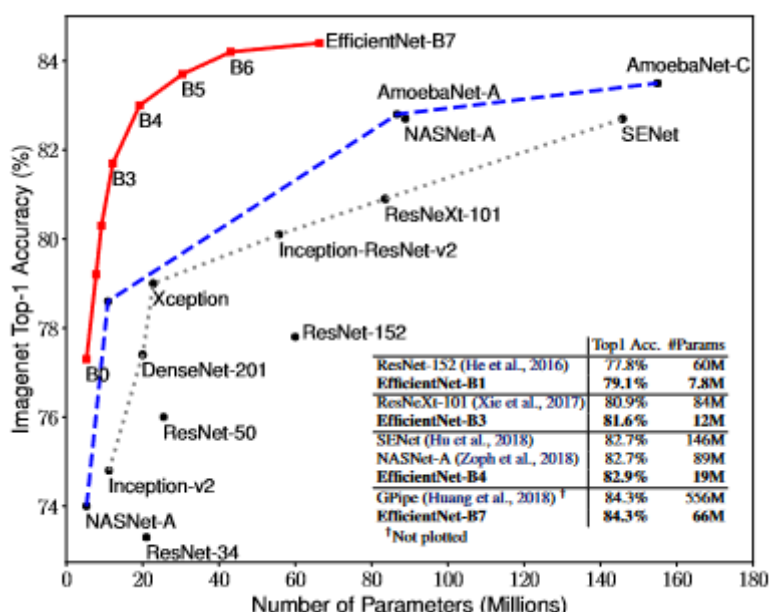
Η εφαρμογή της τεχνητής νοημοσύνης και η ειδικότερα η χρήση των DNN στην ταξινόμηση αλληλουχιών πρωτεϊνών δείχνει μεγάλες δυνατότητες. Τα περισσότερα διαδικτυακά εργαλεία

βασίζονται σε προσεγγίσεις εκμάθησης SVM. Χαρακτηριστικό παράδειγμα είναι το εργαλείο πρόβλεψης προστατευτικών αντιγόνων Vaxijen, του οποίου η AUC της καμπύλης ROC μπορεί να φτάσει μόνο στο 0.743, το οποίο θέλει όριο σε προσεγγίσεις υψηλής ακρίβειας. Επίσης το συνολικό των δεδομένων του περιορίζεται σε 200 πρωτεΐνες. Επομένως είναι χρονοβόρο και επισφαλές να βασιστούμε σε μοντέλα SVM. Κατά συνέπεια, το προτεινόμενο DeepVacPred είναι και πιο γρήγορο και μπορεί να διαχειριστεί δεδομένα από 700.000 διαφορετικές πρωτεΐνες στο σύνολο των δεδομένων.

Το DeepVacPred κατέληξε σε πολλαπλούς επιτόπους μήκους 649aa κατά του SARS-CoV-2. Περιέχει 11 υπομονάδες με 16 επιτόπους Β-κυττάρων, 82 επιτόπους CTL και 89 επιτόπους HTL. Δείχνει καλή αντιγονικότητα, μεγάλη πληθυσμιακή κάλυψη και αποδεκτές φυσικοχημικές ιδιότητες και δομές, το οποίο σημαίνει ότι μπορεί να υπάρξει συνέχεια αυτών των δομών σε πειραματικό και κλινικό επίπεδο. Παράλληλα, υπάρχει και η δυνατότητα/πλαστικότητα του μοντέλου για τις μελλοντικές μεταλλάξεις του ιού, ιδιότητα χαρακτηριστική των κορωνοϊών.

Στο πλαίσιο σύγκρισης των μεθόδων Τεχνητής Νοημοσύνης, που εφαρμόστηκαν στην παρούσα εργασία, δυστυχώς δεν υπάρχουν συγκριτικές αναλύσεις, που να δείχνουν σύγκριση των διαφορετικών μοντέλων πχ σε σενάριο που χρησιμοποιούν κοινή αρχική βάση δεδομένων προς ανάλυση και μελέτη. Αυτό ίσως αποτελέσει μελλοντικό πεδίο μελέτης.

Ωστόσο, υπάρχει σύγκριση όσο αφορά του EfficientNetB4 και του μοντέλου ResNet (Εικόνα 45) (Tan M, 2019).



Εικόνα 45 Σύγκριση μοντέλων TN.

Ειδικότερα, βλέπουμε τη διαφορά που υπάρχει μεταξύ της αύξησης των παραμέτρων κατά την ανάλυση, και το πώς επηρεάζεται η ακρίβεια των μετρήσεων. Το EfficientNet-B4 που εφαρμόστηκε στην ομάδα Wang et al. εμφανίζει μικρότερο αριθμό παραμέτρων, αλλά διατηρεί ακρίβεια στις μετρήσεις.

Καταλήγουμε, λοιπόν, ότι τα υπολογιστικά εργαλεία της τεχνητής νοημοσύνης, πλέον μπορεί να μας βοηθήσει άμεσα και ριζικά στο τομέα της υγείας. Μας παρέχει (η τεχνητή νοημοσύνη) την ικανότητα της πρόβλεψης, την ικανότητας της παγκόσμιας μελέτης του φαινομένου, άμεσα, γρήγορα και με απαντήσεις που μπορούν να σώσουν ζωές.

Αναφορές

- 3D_U-Net_iq.openngenus.org. (n.d.). *iq.openngenus.org*. Retrieved from iq.openngenus.org:
<https://iq.openngenus.org/3d-unet/>
- Adam. (n.d.). *paperwithcode*. Retrieved from paperwithcode:
<https://paperswithcode.com/method/adam>
- Adhikari SP, M. S.-J. (2020). Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (COVID-19) during the early outbreak period: a scoping review. *Infect Dis Poverty*, p. 9:29.
- Agresti A, C. B. (1998). Approximate is better than “exact” for interval estimation of binomial proportions. *Am Stat.* , pp. 52(2):119–126.
- AH, L. (1999). A proof of the triangle inequality for the Tanimoto distance. *J. Math. Chem.*, pp. 26(1–3), 263–265.
- Ahmad T, E. A.-S. (2016). T-cell epitope mapping for the design of powerful vaccines. . *Anal. Chim Acta*, pp. 6:13–22. doi: 10.1016/j.vacrep.2016.07.002.
- AMSGrad. (n.d.). *paperwithcode*. Retrieved from paperwithcode:
<https://paperswithcode.com/method/amsgrad>
- Anand Vibha, D. M. (2008). Probabilistic Asthma Case Finding: A Noisy OR Reformulation. *AMIA Annu Symp Proc.*, pp. 6–10.
- Baell JB, H. G. (2010). New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* , pp. 53(7), 2719–2740 .
- Brenk R, S. A. (2008). Lessons learnt from assembling screening libraries for drug discovery for neglected diseases. *ChemMedChem*, pp. 3(3), 435.
- Bung N., K. S. (2021, Feb 16). De novo design of new chemical entities for SARS-CoV-2 using artificial intelligence. *Future Med Chem.*, pp. doi: 10.4155/fmc-2020-0262.
- C, H. (2020). Coronavirus puts drug repurposing on the fast track. . *Nat. Biotechnol.*, pp. 38(4), 379–381.

- Calders T, J. S. (2007). Efficient AUC optimization for classification. *Knowl. Discov. Databases.*, pp. doi: 10.1007/978-3-540-74976-9_8.
- Chen Y., G. Y. (2020). Structure analysis of the receptor binding of 2019-nCoV. *Biochem. Biophys. Res. Commun*, p. doi: 10.1016/j.bbrc.2020.02.071.
- Chen Y., L. Q. (2020). Emerging coronaviruses: Genome structure, replication, and pathogenesis. *J. Med. Virol.*, pp. 92:418–423. doi: 10.1002/jmv.25681.
- Chen Z-M, F. J.-F. (2020). Diagnosis and treatment recommendations for pediatric respiratory infection caused by the 2019 novel coronavirus. *World J Pediatr.*, pp. 16:240-246.
- Chilamakuri R., A. S. (2021, Jan 21). COVID-19: Characteristics and Therapeutics. *Cells*, σσ. Feb 10(2):206, doi: 10.3390/cells10020206.
- Chung J, G. C. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *ArXiv*, p. Epub ahead of print.
- Çiçek Ö, A. A. (2016). 3D u-net: learning dense volumetric segmentation from sparse annotation. *arXiv*, p. arXiv:1606.06650.
- Çiçek Ö, A. A. (2016). 3D u-net: learning dense volumetric segmentation from sparse annotation. *arXiv*, p. 1606.06650.
- C-index. (χ.χ.). Ανάκτηση από <https://www.statisticshowto.com/c-statistic/>
- CNN_analyticsvidhya. (n.d.). Retrieved from <https://www.analyticsvidhya.com/blog/2021/05/convolutional-neural-networks-understand-the-basics/>
- CNN_wiki. (n.d.). *wikipedia CNN*. Retrieved from *wikipedia CNN*: https://en.wikipedia.org/wiki/Convolutional_neural_network
- Cross_entropy. (n.d.). *ai.nuhil.net*. Retrieved from *ai.nuhil.net*: <https://ai.nuhil.net/deep-learning/cross-entropy>
- Cui J., L. F. (2019). Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.*, pp. 17:181–192. doi: 10.1038/s41579-018-0118-9.
- Danzi GB, L. M. (2020). Acute pulmonary embolism and COVID-19 pneumonia: a random association? *Eur Heart J.*, pp. 41:1858-1858.

- Doveston RG, T. P. (2015). A unified lead-oriented synthesis of over fifty molecular scaffolds. *Org. Biomol. Chem.*, pp. 13(3), 859–865.
- E., C. (2020). Scores of coronavirus vaccines are in competition—How will scientists choose the best? *Nature*, pp. doi: 10.1038/d41586-020-01247-2.
- El-Manzalawy Y, D. D. (2008). Predicting linear B-cell epitopes using string kernels. *J. Mol. Recognit*, pp. 21:243–255. doi: 10.1002/jmr.893.
- Emini EA, H. J. (1985). Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J. Virol.*, pp. 55:836–839.
- Forest, R. (n.d.). *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/Random_forest
- Gaulton A, B. L. (2012). ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.*, pp. 40(D1), D1100–D1107.
- Gussow A.B., A. N. (2020). Genomic determinants of pathogenicity in SARS-CoV-2 and other human coronaviruses. *Proc. Natl. Acad. Sci. USA.*, pp. 117:15193–15199.
- Harrell FE, J. L. (1996). Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med.* , pp. 15(4):361–387. doi: 10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4.
- He K, Z. X. (2015). Deep residual learning for image recognition. . *arXiv*, p. 1512.03385.
- Heagerty PJ, Z. Y. (2005). Survival model predictive accuracy and ROC curves. *Biometrics.*, pp. 61(1):92–105. doi: 10.1111/j.0006-341X.2005.030814.x.
- Heinson A, Y. G. (2017). Enhancing the biological relevance of machine learning classifiers for reverse vaccinology. *Int. J. Mol. Sci.* , p. 18:312. doi: 10.3390/ijms18020312.
- Hellberg S, S. M. (1987). Peptide quantitative structure-activity relationships, a multivariate approach. *Am. Chem. Soc.*, pp. 30:1126–1135. doi: 10.1021/jm00390a003.
- Hobman T.C., M. A. (2020). COVID-19 pandemic: Insights into structure, function, and hACE2 receptor recognition by SARS-CoV-2. . *PLoS Pathog.*, p. 16 doi: 10.1371/journal.ppat.1008762.

- Hoffmann M., K.-W. H. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell.*, pp. 181:271–280. doi: 10.1016/j.cell.2020.02.052.
- IBM_DeepLearning. (n.d.). *IBM*. Retrieved from IBM: <https://www.ibm.com/topics/deep-learning>
- IBM_Random. (n.d.). *IBM*. Retrieved from <https://www.ibm.com/id-en/topics/random-forest>
- IBM_RNN. (n.d.). *IBM*. Retrieved from <https://www.ibm.com/topics/recurrent-neural-networks>
- Igi_global_noisy-or. (n.d.). *Igi global*. Retrieved from What is Noisy-OR model: <https://www.igi-global.com/dictionary/infering-genetic-regulatory-interactions-bayesian/20407>
- Jespersen M, P. B. (2017). BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res.*, pp. 45:24–29 doi: 10.1093/nar/gkx346.
- J-J Z, D. X. (2020). Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China. *Allergy.*, pp. 75:1730-1741.
- Joulin A, M. T. (2015). Inferring algorithmic patterns with stack-augmented recurrent nets. *Presented at: 29th Advances in Neural Information Processing Systems*, (pp. 190–198). Montréal, Canada.
- Jurtz V, e. a. (2017). NetMHCpan-40: Improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J. Immunol.*, pp. 199:3360–3368. doi: 10.4049/jimmunol.1700893.
- Kim D., L. J. (2020). The Architecture of SARS-CoV-2 Transcriptome. *Cell*, pp. 181:914–921.e10. doi: 10.1016/j.cell.2020.04.011.
- Kingma DP, B. J. (2014). Adam: a method for stochastic optimization. *ArXiv*, p. Epub ahead of print.
- Korb, C. R. (2002, August). *Causal Interaction in Bayesian Networks*. Retrieved from ResearchGate:

https://www.researchgate.net/publication/2842592_Causal_Interaction_in_Bayesian_Networks#fullTextFileContent

- Lauer SA, G. K. (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann Intern Med*, pp. 172:577-582.
- Li F. (2016). Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu. Rev. Virol*, pp. 3:237–261. doi: 10.1146/annurev-virology-110615-042301.
- Liu Y., L. C. (2020). The development of Coronavirus 3C-Like protease (3CL(pro)) inhibitors from 2010 to 2020. *Eur. J. Med. Chem.*, p. 206:112711. .
- Loomba RS, V. E. (2020). Covid-19 and Kawasaki syndrome: should we really be surprised? *Cardiol Young*, pp. 30(7):1059-1060.
- Mahmoud I.S., J. Y. (2020). SARS-CoV-2 entry in host cells-multiple targets for treatment and prevention. *Biochimie*, pp. 175:93–98. doi: 10.1016/j.biochi.2020.05.012.
- Mao L, J. H. (2020). Neurologic manifestations of hospitalized patients with coronavirus disease 2019 in Wuhan. *China JAMA Neurol.*, pp. 77:683-690.
- McGuffin LJ, B. K. (2000). The PSIPRED protein structure prediction server. *Bioinformatics.*, pp. 16:1511–1522. doi: 10.1093/bioinformatics/16.4.404.
- Mercurio I., T. V. (2020). Protein structure analysis of the interactions between SARS-CoV-2 spike protein and the human ACE2 receptor: From conformational changes to novel neutralizing antibodies. *Cell Mol. Life Sci.*, pp. doi: 10.1007/s00018-020-03580-1.
- Milletari F, N. N. (2016). V-Net: fully convolutional neural networks for volumetric medical image segmentation. *arXiv*, p. 1606.04797v1.
- Mohamadian M., C. H. (2021, Jan 6). COVID-19: Virology, biology and novel laboratory diagnosis. *J Gene Med.*, pp. 23(2): e3303 , doi: 10.1002/jgm.3303.
- Mollaei H.R., A. A.-N. (2020). Comparison five primer sets from different genome region of COVID-19 for detection of virus infection by conventional RT-PCR. *Iran. J. Microbiol.*, pp. 12:185–193. doi: 10.18502/ijm.v12i3.3234.
- Mousavizadeh L., G. S. (2020). Genotype and phenotype of COVID-19: Their roles in pathogenesis. *J. Microbiol. Immunol. Infect.*, p. doi: 10.1016/j.jmii.2020.03.022.

- Nielsen M, L. C. (2007). NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PLoS ONE*, p. 2:e796. doi: 10.1371/journal.pone.0000796.
- O., A. (1995). The Cartesian product algorithm. . *9th Eur. Conf.* , pp. doi: 10.1007/3-540-49538-X_2.
- Oniško A., D. M. (2001). Learning Bayesian network parameters from small data sets: application of Noisy-OR gates. *Int J Approx Reason*, pp. 27(2):165–182.
- Paszke A, G. S. (2019). Pytorch: an imperative style, high-performance deep learning library. Presented at: 32nd Advances in Neural Information Processing Systems. *Vancouver, Canada*, pp. 8026–8037 (2019).
- PT, T. (2019). On the convergence proof of Amsgrad and a new version. *IEEE Access* 7, pp. 61706–61716.
- Purcell A, M. J. (2007). More than one reason to rethink the use of peptides in vaccine design. *Nat. Rev. Drug Discov.*, pp. 6:404–414. doi: 10.1038/nrd2224.
- Reddi J. Sashank, K. S. (2018). On the Convergence of Adam and Beyond. *ICLR 2018*, p. arXiv:1904.09237.
- Rehman S.U., S. L. (2020). Evolutionary Trajectory for the Emergence of Novel Coronavirus SARS-CoV-2. *Pathogens*, p. 9:240. doi: 10.3390/pathogens9030240.
- ResNet_datagen.tech. (n.d.). *Datagen.tech*. Retrieved from Datagen.tech: <https://datagen.tech/guides/computer-vision/resnet-50/>
- ResNet-50_jason-adam. (n.d.). *JasonAdam.github*. Retrieved from JasonAdam.github: <https://jason-adam.github.io/resnet50/>
- Reynisson B, e. a. (2020). Improved prediction of MHC II antigen presentation through integration and motif deconvolution of mass spectrometry MHC eluted ligand data. *J. Proteome Res*, σ. doi: 10.1021/acs.jproteome.9b00874.
- S., R. (2020). Cutaneous manifestations in COVID-19: a first perspective. *J Eur Acad Dermatol Venereol.*, pp. 34(5):e212-e213.
- Saha S, R. G. (2006). Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins*, pp. 65:40–48. doi: 10.1002/prot.21078.

- Salehi S, A. A. (2020). Coronavirus disease 2019 (COVID-19): a systematic review of imaging findings in 919 patients. *Am J Roentgenol.*, pp. 215(1):87-93.
- Sanders W., F. E. (2020). Comparative analysis of coronavirus genomic RNA structure reveals conservation in SARS-like coronaviruses. *BioRxiv.*, p. doi: 10.1101/2020.06.15.153197.
- Schoeman D., F. B. (2019). Coronavirus envelope protein: Current knowledge. *Viol. J.* , pp. 16:69. doi: 10.1186/s12985-019-1182-0.
- Semiengineering_DL. (n.d.). *semiengineering*. Retrieved from semiengineering: <https://semiengineering.com/deep-learning-spreads/>
- Shereen MA, K. S. (2020). COVID-19 infection: origin, transmission, and characteristics of human coronaviruses. *Journal of Advanced Research*, pp. 24:91-98.
- Softmax_wikipedia. (n.d.). *wikipedia*. Retrieved from wikipedia: https://en.wikipedia.org/wiki/Softmax_function
- statisticshowto. (χ.χ.). *statisticshowto*. Ανάκτηση από <https://www.statisticshowto.com/wald-test/>
- SVM_towardsdatascience. (n.d.). Retrieved from <https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>
- Szegedy C, I. S. (2016). Inception-v4, inception-resnet and the impact of residual connections on learning. *arXiv*, p. 1602.07261.
- Tan M, L. Q. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arxiv*, p. arXiv:1905.11946.
- Trott O, O. A. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.*, pp. 31(2), 455–461 .
- Vital., S. (2020). The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) . *China CDC Wkly.*, pp. 2(8):113–122.
- Vnet_towardsdatascience. (n.d.). *towardsdatascience*. Retrieved from towardsdatascience: <https://towardsdatascience.com/review-v-net-volumetric-convolution-biomedical-image-segmentation-aa15dbaea974>

- Wald_test. (n.d.). www.statisticshowto.com/wald-test/. Retrieved from www.statisticshowto.com/wald-test/: <https://www.statisticshowto.com/wald-test/>
- Wang D, H. B. (2020). Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan. *China Jama*, pp. 323:1061-1069.
- Wang R., J. Z. (2021, Jul 5). Artificial intelligence for prediction of COVID-19 progression using CT imaging and clinical data. *Eur Radiol*, pp. doi: 10.1007/s00330-021-08049-8.
- Wang S, L. W. (2016). CoinFold: A web server for protein contact prediction and contact-assisted protein folding. *Nucleic Acids Res.*, pp. 44:W361–W366. doi: 10.1093/nar/gkw307.
- Wang S, L. W. (2016). RaptorX-property: A web server for protein structure property prediction. *Nucleic Acids Res.* , pp. 44:W430–W435. doi: 10.1093/nar/gkw306.
- Wang Sheng, P. J. (2016, Jan 11). Protein Secondary Structure Prediction Using Deep Convolutional Neural Fields. *Sci Rep.* , p. 6: 18962.
- Weininger D, W. A. (1989). SMILES. 2. Algorithm for generation of unique SMILES notation. *J. Chem. Inf. Comput. Sci.* , pp. 29(2), 97–101 (1989).
- Wiederstein M, S. M. (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* , pp. 35:W407–W410. doi: 10.1093/nar/gkm290.
- Wold S, J. J. (1993). DNA and peptide sequences and chemical processes multivariately modeled by principal component analysis and partial least squares projections to latent structures. *Anal Chim Acta*, pp. 277:239–253. doi: 10.1016/0003-2670(93)80437-P.
- Wu W, L. X. (2019, November 26). A deep learning system that generates quantitative CT reports for diagnosing pulmonary tuberculosis. *arXiv*, p. 1910.02285v1.
- Xu H., Z. L. (2020). High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. *Int. J. Oral Sci.*, pp. 12:8. doi: 10.1038/s41368-020-0074-x.
- Xu X., J. X. (2020, 27 Jun). A Deep Learning System to Screen Novel Coronavirus Disease 2019 Pneumonia. *Engineering (Beijing)*, p. doi: 10.1016/j.eng.2020.04.010.

- Xukun Li, Y. Z. (2021). A deep learning system that generates quantitative CT reports for diagnosing pulmonary Tuberculosis. *Applied Intelligence*, σσ. 51, 4082-4093.
- Yang J., Z. Y. (2020). Prevalence of comorbidities in the novel Wuhan coronavirus (COVID-19) infection: a systematic review and meta-analysis. *Int. J. Infect. Dis.*, pp. 94:91–95.
- Yang X., Y. Y. (2020). Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir. Med.*, pp. 8(5):475–481.
- Yang Z., B. P. (2021, Feb 5). An in silico deep learning approach to multi-epitope vaccine design: a SARS-CoV-2 case study. *Sci Rep*, pp. doi: 10.1038/s41598-021-81749-9.
- Yao B, Z. L. (2012). SVMTriP: A method to predict antigenic epitopes using support vector machine to integrate tri-peptide similarity and propensity. *PLoS ONE*. , p. 7:45152. doi: 10.1371/journal.pone.0045152.
- Zhang K, L. X. (2020). Clinically applicable AI system for accurate diagnosis, quantitative measurements and prognosis of COVID-19 pneumonia using computed tomography. . *Cell*, pp. 181:1423–1433. doi: 10.1016/j.cell.2020.04.045.
- Βλαχάβας Ι., Κ. Π. (2011 (Γ' Έκδοση)). *Τεχνητή Νοημοσύνη*. Θεσσαλονίκη: Εκδόσεις Πανεπιστημίου Μακεδονίας.

