

ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ



IONIAN
UNIVERSITY

Πτυχιακή Εργασία

Παραγωγή Κειμένου από Ιατρικές Εικόνες

ΛΙΘΑΡΗΣ ΝΙΚΟΛΑΟΣ

Επιβλέπων: ΚΑΝΑΒΟΣ ΑΝΔΡΕΑΣ

August 25, 2024

Επιβλέπων

ΚΑΝΑΒΟΣ ΑΝΔΡΕΑΣ, Αναπληρωτής Καθηγητής,
IONIO ΠΑΝΕΠΙΣΤΗΜΙΟ

Τριμελής Επιτροπή

ΚΑΝΑΒΟΣ ΑΝΔΡΕΑΣ, Αναπληρωτής Καθηγητής,
IONIO ΠΑΝΕΠΙΣΤΗΜΙΟ

ΚΕΡΜΑΝΙΔΟΥ ΚΑΤΙΑ - ΛΗΔΑ, Αναπληρώτρια Καθηγήτρια,
IONIO ΠΑΝΕΠΙΣΤΗΜΙΟ

ΣΩΤΗΡΟΠΟΥΛΟΥ ΑΝΝΑ, Εργαστηριακό Διδακτικό Προσωπικό,
IONIO ΠΑΝΕΠΙΣΤΗΜΙΟ

Περίληψη

Η περιγραφή ιατρικών εικόνων είναι ένας αναδυόμενος τομέας που βρίσκεται στη διασταύρωση της όρασης υπολογιστών και της υγειονομικής περίθαλψης. Επιδιώκει να αντιμετωπίσει την πολύπλοκη πρόκληση της αυτόματης δημιουργίας περιγραφικών και κλινικά κατάλληλων περιγραφών κειμένου για ιατρικές εικόνες. Η ικανότητα αυτή έχει σημαντικές επιπτώσεις στην ενίσχυση της κατανόησης και της ερμηνείας των ιατρικών εικόνων, βοηθώντας τελικά τους επαγγελματίες υγείας στις διαδικασίες λήψης κλινικών αποφάσεων.

Η παρούσα διατριβή εισάγει μια νέα προσέγγιση για τη σύνταξη υποτίτλων ιατρικών εικόνων αξιοποιώντας έναν συνδυασμό Συνελικτικών Νευρωνικών Δικτύων (CNN) και Αναδρομικών Νευρωνικών Δικτύων (RNN). Το CNN κομμάτι του μοντέλου είναι υπεύθυνο για την εξαγωγή σημαντικών οπτικών χαρακτηριστικών από ιατρικές εικόνες. Αυτά τα χαρακτηριστικά χρησιμοποιούνται στη συνέχεια από το RNN κομμάτι για τη δημιουργία συνεκτικών και κατατοπιστικών υποτίτλων. Η συνεργασία μεταξύ CNN και RNN επιτρέπει στο μοντέλο να συλλαμβάνει και να περιγράφει αποτελεσματικά τις περίπλοκες λεπτομέρειες και τις πολύπλοκες οπτικές αναπαραστάσεις που υπάρχουν στις ιατρικές εικόνες.

Ένας από τους πρωταρχικούς στόχους του προτεινόμενου μοντέλου είναι η βελτίωση της ερμηνευσιμότητας των ιατρικών εικόνων, οι οποίες συχνά χαρακτηρίζονται από πολύπλοκες ανατομικές δομές, παθολογικές ανωμαλίες και λεπτές οπτικές ενδείξεις που είναι ζωτικής σημασίας για την ακριβή διάγνωση. Χρησιμοποιώντας CNN, το μοντέλο μπορεί να μάθει να αναγνωρίζει και να κωδικοποιεί αυτά τα κρίσιμα οπτικά στοιχεία. Τα CNN είναι γνωστά για την ικανότητά τους να εκτελούν εξαγωγή χαρακτηριστικών υψηλού επιπέδου, η οποία είναι ιδιαίτερα επωφελής για το χειρισμό των αποχρώσεων των οπτικών πληροφοριών που περιέχονται στις ιατρικές εικόνες. Για παράδειγμα, η χρήση των CNNs στο "Neural Captioning for the ImageCLEF 2017 Medical Image Challenges" κατέδειξε την αποτελεσματικότητά τους στην εξαγωγή ιεραρχικών χαρακτηριστικών για τη λεζάντα ιατρικών εικόνων.

Το RNN του μοντέλου διαδραματίζει κρίσιμο ρόλο στη δημιουργία υποτίτλων που δεν είναι

μόνο συνεκτικοί αλλά και σχετικοί με το πλαίσιο. Αυτό είναι ιδιαίτερα σημαντικό στον ιατρικό τομέα, όπου οι ακριβείς και αδιαμφισβήτητες περιγραφές είναι απαραίτητες για την αποτελεσματική επικοινωνία και τη λήψη αποφάσεων. Τα RNN είναι ικανά στο χειρισμό διαδοχικών δεδομένων, γεγονός που τα καθιστά κατάλληλα για τη δημιουργία περιγραφών φυσικής γλώσσας με βάση τα χαρακτηριστικά που εξάγονται από τα CNN.

Μια βασική καινοτομία στο προτεινόμενο μοντέλο είναι η ενσωμάτωση μηχανισμών προσοχής. Οι μηχανισμοί προσοχής επιτρέπουν στο μοντέλο να εστιάζει επιλεκτικά στις πιο σχετικές περιοχές της εικόνας εισόδου κατά τη διαδικασία δημιουργίας περιγραφών. Αυτή η ικανότητα είναι κρίσιμη στο πλαίσιο των ιατρικών εικόνων, οι οποίες συχνά περιέχουν πολλαπλές περιοχές ενδιαφέροντος. Με την ιεράρχηση και την προσοχή στις διαγνωστικά σημαντικές περιοχές, το μοντέλο μπορεί να παράγει λεζάντες που είναι πιο συγκεκριμένες και κατατοπιστικές.

Εκτός από τις ποσοτικές αξιολογήσεις, η παρούσα διατριβή περιλαμβάνει επίσης ποιοτικές αναλύσεις και μελέτες περιπτώσεων που καταδεικνύουν την ικανότητα του μοντέλου να παράγει κλινικά σχετικές και κατατοπιστικές λεζάντες. Αυτές οι ποιοτικές αξιολογήσεις αναδεικνύουν τον πιθανό αντίκτυπο της προτεινόμενης προσέγγισης στη βελτίωση των κλινικών ροών εργασίας. Για παράδειγμα, παρέχοντας αυτόματες περιγραφές ιατρικών εικόνων, το μοντέλο μπορεί να βοηθήσει τους ακτινολόγους και άλλους επαγγελματίες υγείας στις διαγνωστικές τους διαδικασίες, μειώνοντας ενδεχομένως τον χρόνο που απαιτείται για την ερμηνεία των εικόνων και βελτιώνοντας τη διαγνωστική ακρίβεια.

Οι ποιοτικές αναλύσεις υπογραμμίζουν επίσης την ικανότητα του μοντέλου να παράγει λεζάντες που διευκολύνουν την καλύτερη επικοινωνία μεταξύ των επαγγελματιών υγείας. Στις κλινικές ρυθμίσεις, οι σαφείς και συνοπτικές περιγραφές των ιατρικών εικόνων είναι απαραίτητες για την αποτελεσματική συνεργασία και τη λήψη αποφάσεων. Η ικανότητα του προτεινόμενου μοντέλου να παράγει λεζάντες υψηλής ποιότητας μπορεί να βοηθήσει στην προετοιμασία ιατρικών αναφορών, εκπαιδευτικού υλικού και τεκμηρίωσης, συμβάλλοντας έτσι στη βελτίωση της φροντίδας και των αποτελεσμάτων των ασθενών.

Η διατριβή διερευνά επίσης τις πιθανές εφαρμογές του προτεινόμενου μοντέλου πέραν της ακτινολογίας και της παθολογίας. Για παράδειγμα, το μοντέλο θα μπορούσε να προσαρμοστεί σε άλλες μεθόδους ιατρικής απεικόνισης, όπως η μαγνητική τομογραφία ή η αξονική τομογραφία, ώστε να παράγει λεζάντες για ένα ευρύτερο φάσμα ιατρικών εικόνων. Επιπλέον, η ενσωμάτωση του μοντέλου σε συστήματα υποστήριξης κλινικών αποφάσεων θα μπορούσε να ενισχύσει περαιτέρω τη χρησιμότητά του, παρέχοντας βοήθεια σε πραγματικό χρόνο στους επαγγελματίες υγείας κατά τη διάρκεια της διαγνωστικής διαδικασίας.

Πρόλογος και Ευχαριστίες

Αυτή η ερευνητική εργασία είναι το αποτέλεσμα ενός υπέροχου ταξιδιού γεμάτο μάθηση, ανακαλύψεις και προσωπική ανάπτυξη. Πρώτα απ' όλα, θα ήθελα να ευχαριστήσω θερμά τον καθηγητή μου, Δρ. Ανδρέα Καναβό. Η συνεχής του υποστήριξη, οι συμβουλές και η αφοσίωσή σας ήταν ανεκτίμητες. Σας ευχαριστώ ειλικρινά για την υπομονή σας και για τον χρόνο που αφιερώσατε στην ανάπτυξή μου. Θα ήθελα επίσης να ευχαριστήσω θερμά το Ιόνιο Πανεπιστήμιο. Το ίδρυμα αυτό μου προσέφερε ένα υποστηρικτικό περιβάλλον που ενθαρρύνει την πνευματική περιέργεια και την καινοτομία. Οι πόροι, οι εγκαταστάσεις και η ακαδημαϊκή κοινότητα υπήρξαν πολύτιμοι για την υποστήριξη των ερευνητικών μου προσπαθειών. Τέλος, θέλω να εκφράσω την ευγνωμοσύνη μου στην οικογένεια και στους φίλους μου για την αδιάλειπτη υποστήριξη, ενθάρρυνση και κατανόηση καθ' όλη τη διάρκεια αυτής της διαδρομής. Η πίστη σας σε εμένα υπήρξε σταθερή πηγή κινήτρου.

Περιεχόμενα

A	Εισαγωγή	1
A.1	Η ΣΗΜΑΣΙΑ ΤΗΣ ΙΑΤΡΙΚΗΣ ΑΠΕΙΚΟΝΙΣΗΣ	1
A.2	ΠΡΟΒΛΗΜΑ	2
A.3	ΣΤΟΧΟΙ	3
B	Ανασκόπηση βιβλιογραφίας	5
B.1	ΥΠΟΤΙΤΛΟΙ ΕΙΚΟΝΑΣ ΣΤΗΝ ΟΡΑΣΗ ΥΠΟΛΟΓΙΣΤΩΝ	11
B.1.1	Πρώιμες προσεγγίσεις: Μέθοδοι βασισμένες σε πρότυπα	11
B.1.2	Εξελιξείς στη βαθιά μάθηση	12
B.1.3	Μετρικές αξιολόγησης	16
B.1.4	Προκλήσεις και μελλοντικές κατευθύνσεις	16
B.2	CNN ΣΤΗΝ ΙΑΤΡΙΚΗ ΑΠΕΙΚΟΝΙΣΗ	17
B.2.1	Sequential Data Processing	23
B.2.2	Δημιουργία κειμένου σε ιατρικές εικόνες με λεζάντες	23
B.2.3	Μηχανισμοί προσοχής	24
B.2.4	Εκπαίδευση RNN για τη λεζάντα ιατρικών εικόνων	25
B.2.5	Προκλήσεις και μελλοντικές κατευθύνσεις	25
B.3	ΜΗΧΑΝΙΣΜΟΙ ΠΡΟΣΟΧΗΣ	26
B.3.1	Προέλευση και ανάπτυξη	27
B.3.2	Τύποι μηχανισμών προσοχής	27
B.3.3	Εφαρμογή στον υποτιτλισμό ιατρικών εικόνων	28
B.3.4	Οφέλη και αντίκτυπος	29
B.4	ΥΒΡΙΔΙΚΕΣ ΑΡΧΙΤΕΚΤΟΝΙΚΕΣ CNN-RNN	30
B.5	ΠΡΟΗΓΟΥΜΕΝΕΣ ΜΕΛΕΤΕΣ	30

B.6	ΜΟΝΤΕΛΑ ΠΡΟΣΟΧΗΣ	31
B.7	ΜΕΤΑΣΧΗΜΑΤΙΣΤΕΣ	31
B.8	ΠΡΟΚΛΗΣΕΙΣ ΚΑΙ ΜΕΛΛΟΝΤΙΚΕΣ ΚΑΤΕΥΘΥΝΣΕΙΣ	33
C	Μεθοδολογία	35
C.1	ΕΠΙΣΚΟΠΗΣΗ	35
C.2	ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ	35
C.3	ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ	36
C.3.1	Προεπεξεργασία εικόνας	36
C.3.2	Προεπεξεργασία κειμένου	36
C.4	ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΜΟΝΤΕΛΟΥ	37
C.4.1	Encoder - Image Feature Model	37
C.4.2	Encoder - Sequence Feature Model	38
C.4.3	Attention Mechanism	38
C.4.4	Decoder	38
C.4.5	Model Compilation	39
C.4.6	Callbacks	39
C.5	CNN ΓΙΑ ΕΞΑΓΩΓΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ	40
C.6	RNN ΓΙΑ ΤΗ ΔΗΜΙΟΥΡΓΙΑ ΛΕΞΑΝΤΑΣ	40
C.7	ΔΙΑΔΙΚΑΣΙΑ ΕΚΠΑΙΔΕΥΣΗΣ	42
C.8	ΜΕΤΡΙΚΕΣ ΑΞΙΟΛΟΓΗΣΗΣ	43
C.9	ΠΕΙΡΑΜΑΤΙΚΗ ΡΥΘΜΙΣΗ	43
D	Αποτελέσματα	45
D.1	ΠΟΣΟΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ	45
D.1.1	Παραδοσιακές Μετρικές	45
D.1.2	Μετρική BLUE	47
D.1.3	Βαθμολογία BLEU-1	48
D.1.4	Βαθμολογία BLEU-2	48
D.2	ΠΟΙΟΤΙΚΗ ΑΝΑΛΥΣΗ	49
D.2.1	Παράδειγμα 1	50
D.2.2	Παράδειγμα 2	51
D.2.3	Παράδειγμα 3	51

D.2.4	Παράδειγμα 4	52
D.2.5	Παράδειγμα 5	53
D.3	ΣΥΓΚΡΙΤΙΚΗ ΑΝΑΛΥΣΗ	53
D.4	ΑΝΑΛΥΣΗ ΑΠΟΔΟΣΗΣ	54
D.5	ΑΝΑΛΥΣΗ ΕΠΙΔΟΣΕΩΝ	54
D.6	ΕΠΠΤΩΣΕΙΣ ΓΙΑ ΤΗΝ ΚΛΙΝΙΚΗ ΠΡΑΚΤΙΚΗ	55
E	Συζήτηση	57
E.1	ΕΡΜΗΝΕΙΑ ΤΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ	57
E.2	ΣΗΜΑΣΙΑ ΤΩΝ ΥΨΗΛΩΝ ΒΑΘΜΟΛΟΓΙΩΝ BLEU	57
E.3	ΕΠΙΚΥΡΩΣΗ ΤΗΣ ΧΡΗΣΙΜΟΤΗΤΑΣ ΤΟΥ ΜΟΝΤΕΛΟΥ	58
E.4	ΕΥΡΥΤΕΡΕΣ ΕΠΠΤΩΣΕΙΣ	59
E.5	ΜΕΛΛΟΝΤΙΚΕΣ ΚΑΤΕΥΘΥΝΣΕΙΣ	60
E.6	ΠΕΡΙΟΡΙΣΜΟΙ ΤΟΥ ΜΟΝΤΕΛΟΥ	60
E.7	ΣΥΝΕΠΕΙΕΣ	63
E.8	ΣΥΜΠΕΡΑΣΜΑ	65
F	Συμπεράσματα	66
F.1	ΣΥΝΟΨΗ ΤΩΝ ΣΥΝΕΙΣΦΟΡΩΝ	66
F.2	ΤΕΛΙΚΕΣ ΣΚΕΨΕΙΣ	68
F.3	ΕΝΔΥΝΑΜΩΣΗ ΤΩΝ ΕΠΑΓΓΕΛΜΑΤΙΩΝ ΤΗΣ ΥΓΕΙΟΝΟΜΙΚΗΣ ΠΕΡΙΘΑΛΨΗΣ	71
F.4	ΠΡΩΘΗΣΗ ΤΗΣ ΔΙΕΠΙΣΤΗΜΟΝΙΚΗΣ ΣΥΝΕΡΓΑΣΙΑΣ ΚΑΙ ΤΗΣ ΚΑΙΝΟΤΟΜΙΑΣ	71
F.5	ΜΕΤΑΣΧΗΜΑΤΙΣΤΙΚΕΣ ΑΛΛΑΓΕΣ ΣΤΗΝ ΥΓΕΙΟΝΟΜΙΚΗ ΠΕΡΙΘΑΛΨΗ	72
	Συντμήσεις	77
	Γλωσσάρι Ξενικών Όρων	79
	Κώδικας	80

Κατάλογος Σχημάτων

D.1	Chest x-ray showing no foci in the lungs.	50
D.2	Abdominal ctscan image suggest with simple cyst in liver segment.	51
D.3	anteroposterior and lateral radiograph of shoulder depicts soft tissue tumor.	51
D.4	X-ray showing asymmetry of clavicle due to right sided dislocation.	52
D.5	Reformatted ctscan showing volar ulnar and dorsal sides of wrist.	53

Κατάλογος Πινάκων

<i>C.1</i>	<i>Λεπτομέρειες Συνόλου Δεδομένων</i>	36
<i>D.1</i>	<i>Μετρικές απόδοσης</i>	46
<i>D.2</i>	<i>Αποτελέσματα αξιολόγησης του μοντέλου</i>	48
<i>D.3</i>	<i>Συγκριτική ανάλυση των αλγορίθμων με τις μετρικές BLUE-1 & BLUE-2</i>	54
<i>F.1</i>	<i>Model Summary: "OLIVER"</i>	88

Κεφάλαιο Α

Εισαγωγή

Α.1 Η σημασία της ιατρικής απεικόνισης

Η ιατρική απεικόνιση διαδραματίζει καθοριστικό ρόλο στη σύγχρονη υγειονομική περίθαλψη, παρέχοντας στους γιατρούς μη-επεμβατικές γνώσεις για τη δομή και τη λειτουργία του ανθρώπινου σώματος. Μέθοδοι όπως οι ακτίνες X, η μαγνητική τομογραφία (MRI), η αξονική/υπολογιστική τομογραφία (CT) και οι ιστοπαθολογικές διαφάνειες έχουν γίνει απαραίτητα εργαλεία για τη διάγνωση και την παρακολούθηση ενός ευρέος φάσματος ιατρικών καταστάσεων. Ο πολλαπλασιασμός αυτών των μεθόδων απεικόνισης έχει οδηγήσει σε έναν συνεχώς αυξανόμενο όγκο ψηφιακών ιατρικών εικόνων που απαιτούν ερμηνεία και ανάλυση. Η πολυπλοκότητα και η ποικιλομορφία αυτών των εικόνων απαιτούν προηγμένα υπολογιστικά εργαλεία για να βοηθήσουν τους επαγγελματίες υγείας να κάνουν ακριβείς διαγνώσεις. Η ενσωμάτωση τεχνικών Artificial Intelligence (AI) και Deep Learning (DL) στην ιατρική απεικόνιση μεταμορφώνει το τοπίο της διαγνωστικής απεικόνισης, προσφέροντας νέες δυνατότητες για βελτιωμένα αποτελέσματα για τους ασθενείς [2, 10].

Η πολυπλοκότητα και ο πλούτος των ιατρικών εικόνων θέτουν σημαντικές προκλήσεις για τους επαγγελματίες του τομέα της υγείας, απαιτώντας συχνά εξειδικευμένη εκπαίδευση και εμπειρογνωμοσύνη για ακριβή διάγνωση και σχεδιασμό θεραπείας. Για παράδειγμα, οι ακτινολόγοι πρέπει να εξετάζουν σχολαστικά τις αξονικές τομογραφίες για να εντοπίζουν ανεπαίσθητα σημάδια νόσου, όπως μικρές βλάβες που μπορεί να υποδηλώνουν καρκίνο σε πρώιμο στάδιο. Οι παθολόγοι, από την άλλη πλευρά, αναλύουν ιστοπαθολογικές αντικειμενοφόρους πλάκες για να εντοπίσουν κυτταρικές ανωμαλίες, οι οποίες είναι ζωτικής σημασίας για τη διάγνωση

καταστάσεων όπως ο καρκίνος και οι φλεγμονώδεις νόσοι [13, 11].

Οι εξελίξεις στην τεχνολογία απεικόνισης έχουν οδηγήσει σε αυξημένη ανάλυση και λεπτομέρεια, επιτρέποντας ακριβέστερες διαγνώσεις. Ωστόσο, αυτό σημαίνει επίσης ότι ο όγκος των δεδομένων που παράγονται είναι τεράστιος, οδηγώντας σε πιθανή υπερφόρτωση πληροφοριών για τους κλινικούς γιατρούς. Η ψηφιακή βάση δεδομένων για τη μαστογραφία προσυμπτωματικού ελέγχου (DDSM) περιέχει χιλιάδες μαστογραφίες, καθεμία από τις οποίες απαιτεί προσεκτική ανάλυση για τον εντοπισμό ενδείξεων καρκίνου του μαστού, όπως μικροασβεστώσεις και μάζες. Ομοίως, το σύνολο δεδομένων ISIC για εικόνες δερματικών βλαβών παρέχει μια τεράστια συλλογή δερματοσκοπικών εικόνων που χρησιμοποιούνται για την ανίχνευση μελανώματος, αναδεικνύοντας την ανάγκη για αποτελεσματικά εργαλεία ανάλυσης εικόνων [21].

Παρά τη διαθεσιμότητα δεδομένων απεικόνισης υψηλής ποιότητας, η χειροκίνητη ερμηνεία ιατρικών εικόνων είναι χρονοβόρα και υπόκειται σε ανθρώπινα σφάλματα. Η αυξανόμενη ζήτηση για υπηρεσίες ιατρικής απεικόνισης, σε συνδυασμό με την έλλειψη εκπαιδευμένων ακτινολόγων και παθολόγων, επιδεινώνει αυτές τις προκλήσεις. Αυτό υπογραμμίζει την επείγουσα ανάγκη για αυτοματοποιημένα εργαλεία που μπορούν να βοηθήσουν στην ερμηνεία των ιατρικών εικόνων, εξασφαλίζοντας έγκαιρες και ακριβείς διαγνώσεις [7].

A.2 Πρόβλημα

Τα τελευταία χρόνια, η ανάπτυξη της τεχνητής νοημοσύνης (AI) και της βαθιάς μάθησης (Deep Learning) καινοτομεί αρκετά στην ανάλυση ιατρικών εικόνων, με τη δυνατότητα να ενισχύσει τις ανθρώπινες ικανότητες και να βελτιώσει τις κλινικές ροές εργασίας. Οι τεχνικές τεχνητής νοημοσύνης, ιδίως εκείνες που περιλαμβάνουν βαθιά μάθηση, έχουν δείξει μεγάλη υπόσχεση για την αυτοματοποίηση διαφόρων εργασιών, όπως η ταξινόμηση εικόνων, η ανίχνευση αντικειμένων και η κατάτμηση [5]. Αυτές οι εξελίξεις έχουν θέσει τις βάσεις για πιο σύνθετες εφαρμογές, συμπεριλαμβανομένου του υποτιτλισμού ιατρικών εικόνων, η οποία αποσκοπεί στην αυτόματη δημιουργία περιγραφικών και κλινικά σχετικών περιλήψεων κειμένου για ιατρικές εικόνες [17].

Η περιγραφή ιατρικών εικόνων γεφυρώνει το χάσμα μεταξύ της ανάλυσης εικόνας και της επεξεργασίας φυσικής γλώσσας (Natural Language Processing - NLP), επιτρέποντας τη δημιουργία κειμενικών περιγραφών που μπορούν να μεταφέρουν κρίσιμες πληροφορίες σχετικά

με το οπτικό περιεχόμενο. Παρέχοντας ερμηνεύσιμες και πλούσιες σε περιεχόμενο περιγραφές, οι υπότιτλοι ιατρικών εικόνων έχουν τη δυνατότητα να βελτιώσουν την παροχή υγειονομικής περίθαλψης, την ιατρική εκπαίδευση και την έρευνα [14, 22].

Ωστόσο, παραμένουν σημαντικές προκλήσεις στην ανάπτυξη ισχυρών μοντέλων υποτιτλισμού ιατρικών εικόνων. Τα υπάρχοντα μοντέλα συχνά δυσκολεύονται να αντιμετωπίσουν την ποικιλόμορφη και πολύπλοκη φύση των ιατρικών εικόνων, με αποτέλεσμα να δημιουργούνται προβλήματα γενικευσιμότητας και ακρίβειας. Για παράδειγμα, τα μοντέλα που εκπαιδεύτηκαν σε ακτινογραφίες θώρακος από το σύνολο δεδομένων του Indiana University Chest X-Ray μπορεί να μην αποδίδουν το ίδιο καλά σε εικόνες από διαφορετικούς πληθυσμούς ασθενών ή συνθήκες απεικόνισης [4]. Οι ανακριβείς ή μη-ειδικές λεζάντες μπορεί να οδηγήσουν σε λανθασμένες διαγνώσεις και ακατάλληλες αποφάσεις θεραπείας. Ως εκ τούτου, είναι ζωτικής σημασίας η ανάπτυξη μοντέλων που όχι μόνο παράγουν ακριβείς περιγραφές αλλά και αποτυπώνουν τις αποχρώσεις λεπτομέρειες που είναι κρίσιμες για τη λήψη κλινικών αποφάσεων [10].

A.3 Στόχοι

Ο πρωταρχικός στόχος της παρούσας ερευνητικής εργασίας είναι η ανάπτυξη ενός αξιόπιστου και αποδοτικού μοντέλου για την παραγωγή υποτίτλων ιατρικών εικόνων που αξιοποιεί τα πλεονεκτήματα τόσο των Convolutional Neural Networks (CNN) όσο και των Recurrent Neural Networks (RNN). Συγκεκριμένα, η έρευνα στοχεύει στα εξής:

- **Ανάπτυξη υβριδικής αρχιτεκτονικής CNN-RNN:** Σχεδιασμός ενός μοντέλου που ενσωματώνει CNN για την εξαγωγή οπτικών χαρακτηριστικών με RNN για τη διαδοχική παραγωγή κειμένου, εξασφαλίζοντας ότι το μοντέλο μπορεί να χειριστεί τις περίπλοκες λεπτομέρειες που υπάρχουν στις ιατρικές εικόνες και να παράγει συνεκτικούς και σχετικούς με το περιεχόμενο υποτίτλους [16, 8]. Το CNN θα επικεντρωθεί στην εξαγωγή ιεραρχικών οπτικών χαρακτηριστικών, καταγράφοντας τόσο λεπτομέρειες χαμηλού επιπέδου (όπως υφές και ακμές) όσο και σημασιολογία υψηλού επιπέδου (όπως ανατομικές δομές και παθολογικά ευρήματα). Το RNN, ιδίως με αμφίδρομα στρώματα LSTM, θα παράγει συνεκτικές και κατάλληλες από άποψη πλαισίου λεζάντες με βάση αυτά τα εξαγόμενα χαρακτηριστικά.
- **Ενσωμάτωση μηχανισμών προσοχής:** Επιπλέον, ενισχύθηκε η ικανότητα του

μοντέλου να εστιάζει στις πιο σχετικές περιοχές της εικόνας ενσωματώνοντας μηχανισμούς προσοχής, οι οποίοι είναι εξαιρετικής σημασίας για την ακριβή περιγραφή διαγνωστικά σημαντικών χαρακτηριστικών [23, 25]. Ο μηχανισμός προσοχής θα επιτρέψει στο μοντέλο να σταθμίζει δυναμικά διάφορα μέρη της εικόνας, διασφαλίζοντας ότι οι παραγόμενες λεζάντες δίνουν έμφαση σε κλινικά σημαντικές περιοχές, όπως περιοχές με ανωμαλίες ή βασικά ανατομικά ορόσημα.

- **Αξιολόγηση της απόδοσης του μοντέλου:** Η απόδοση του μοντέλου αξιολογήθηκε χρησιμοποιώντας τόσο ποσοτικές μετρικές όπως το BLUE όσο και ποιοτική ανάλυση, συγκρίνοντας το με υπάρχοντα μοντέλα τελευταίας τεχνολογίας για ναδειχθεί η αποτελεσματικότητά του [26]. Η αξιολόγηση περιλαμβάνει ένα ολοκληρωμένο σύνολο μετρικών που μετρούν την ακρίβεια, τη συνοχή και την κλινική συνάφεια των παραγόμενων υποτίτλων.

Κεφάλαιο Β

Ανασκόπηση βιβλιογραφίας

Η ανάλυση ιατρικών εικόνων έχει μετασχηματίσει την υγειονομική περίθαλψη, παρέχοντας σημαντική κατανόηση για τη δομή και τη λειτουργία του ανθρώπινου σώματος. Ο τομέας αυτός περιλαμβάνει διάφορες μεθόδους απεικόνισης, καθεμία από τις οποίες έχει ξεχωριστές δυνατότητες και εφαρμογές που ενισχύουν την κλινική πρακτική και την ιατρική έρευνα. Οι ακτίνες Χ παρέχουν γρήγορη και αποτελεσματική απεικόνιση σκελετικών δομών και ασθενειών, ενώ η μαγνητική τομογραφία (MRI) προσφέρει λεπτομερή εικόνα των μαλακών ιστών και του κεντρικού νευρικού συστήματος. Η υπολογιστική τομογραφία (CT) είναι καίρια για την απεικόνιση πολύπλοκων δομών όπως τον εγκέφαλο και τις αγγειακές παθήσεις, ενώ οι ιστοπαθολογικές διαφάνειες παρέχουν ακριβείς πληροφορίες για την κατανόηση κυτταρικών και ιστικών διαταραχών. Η ανάπτυξη και η εξέλιξη αυτών των τεχνολογιών συνεχώς βελτιώνει τη διάγνωση, τη θεραπεία και την επιστημονική έρευνα στον τομέα της ιατρικής [3].

Με την πρόοδο της τεχνητής νοημοσύνης, ειδικά με τη χρήση προηγμένων μοντέλων μηχανικής μάθησης, όπως τα βαθιά συνελικτικά νευρωνικά δίκτυα (CNN) και τα αναδρομικά νευρωνικά δίκτυα (RNN), η ιατρική απεικόνιση έχει περάσει σε μια νέα εποχή [20]. Τα μοντέλα αυτά επιτρέπουν την αυτόματη ανάλυση και ερμηνεία ιατρικών εικόνων, προσφέροντας ακριβείς διαγνώσεις και προβλέψεις. Για παράδειγμα, η χρήση CNNs για την αναγνώριση μοτίβων σε ακτινογραφίες μπορεί να βοηθήσει στη γρήγορη ανίχνευση καταγμάτων ή άλλων ανωμαλιών [20]. Αντίστοιχα, τα RNNs μπορούν να χρησιμοποιηθούν για την ανάλυση σειρών εικόνων, όπως η παρακολούθηση της εξέλιξης μιας ασθένειας μέσω διαδοχικών MRI [3].

Επιπλέον, η σύγχρονη έρευνα επικεντρώνεται στη βελτίωση της ακρίβειας και της ταχύτητας των μεθόδων αυτών μέσω της εκπαίδευσης των μοντέλων σε μεγαλύτερα και πιο ποικίλα

σύνολα δεδομένων. Η ενσωμάτωση τεχνικών όπως η μεταφορά μάθησης επιτρέπει τη χρήση προεκπαιδευμένων μοντέλων, τα οποία μπορούν να προσαρμοστούν σε συγκεκριμένα ιατρικά δεδομένα με μικρή επιπλέον εκπαίδευση [20]. Η δυνατότητα αυτή ανοίγει νέους ορίζοντες στην εξατομικευμένη ιατρική, καθώς επιτρέπει την ανάπτυξη προσαρμοσμένων διαγνωστικών εργαλείων για συγκεκριμένες ασθένειες ή πληθυσμούς [3].

Συνολικά, η ανάλυση ιατρικών εικόνων με τη χρήση προηγμένων τεχνικών τεχνητής νοημοσύνης και μηχανικής μάθησης αναμένεται να φέρει σημαντικές βελτιώσεις στην ποιότητα της υγειονομικής περίθαλψης, προσφέροντας ταχύτερες και ακριβέστερες διαγνώσεις, καλύτερη παρακολούθηση της θεραπείας και νέες δυνατότητες στην ιατρική έρευνα [20].

- **Απεικόνιση με ακτίνες X:** Η απεικόνιση με ακτίνες X είναι μία από τις παλαιότερες και πιο ευρέως χρησιμοποιούμενες τεχνικές στην ιατρική διάγνωση. Περιλαμβάνει τη διέλευση ακτίνων X μέσα από το σώμα για τη λήψη εικόνων των εσωτερικών δομών, αποδεικνύοντας την αποτελεσματικότητά της στην απεικόνιση των οστών και την ανίχνευση καταγμάτων, λοιμώξεων και άλλων ανωμαλιών [1, 13]. Παρά την ευρεία της χρήση, η ερμηνεία των εικόνων ακτίνων X μπορεί να είναι προβληματική λόγω των επικαλυπτόμενων δομών και της ανάγκης για υψηλή ανάλυση αντίθεσης για τη διάκριση μεταξύ διαφορετικών ιστών. Για παράδειγμα, η διαφοροποίηση μεταξύ μαλακών ιστών και οστών ή η ανίχνευση μικρών καταγμάτων μπορεί να απαιτεί προηγμένες τεχνικές επεξεργασίας εικόνας για την επίτευξη ακριβούς διάγνωσης.

Η απεικόνιση με ακτίνες X χρησιμοποιείται επίσης εκτενώς σε εξετάσεις θώρακα για τη διάγνωση καταστάσεων όπως η πνευμονία, η φυματίωση και ο καρκίνος του πνεύμονα, γεγονός που την καθιστά κρίσιμο εργαλείο τόσο σε επείγουσες καταστάσεις όσο και σε ιατρικές καταστάσεις ρουτίνας [1, 13]. Επιπλέον, προηγμένες τεχνικές όπως η απορροφησιμετρία ακτίνων X διπλής ενέργειας (DEXA) έχουν επιτρέψει την αξιολόγηση της οστικής πυκνότητας, η οποία είναι ζωτικής σημασίας για τη διάγνωση της οστεοπόρωσης και άλλων μεταβολικών ασθενειών των οστών.

- **Απεικόνιση μαγνητικού συντονισμού (MRI):** Η μαγνητική τομογραφία (MRI) αποτελεί μια προηγμένη τεχνική απεικόνισης που εκμεταλλεύεται ισχυρά μαγνητικά πεδία και ραδιοκύματα για την αναπαράσταση λεπτομερών εικόνων των εσωτερικών δομών του ανθρώπινου σώματος. Η ευρεία της χρήση προκύπτει από την ικανότητά της να προσφέρει εξαιρετική αντίθεση μεταξύ διαφορετικών τύπων ιστών, καθιστώντας την ιδανική για τη διάγνωση ποικίλων καταστάσεων όπως όγκοι, εγκεφαλικά επεισόδια και μυ-

οσκελετικές διαταραχές [9]. Η ακρίβεια και η λεπτομέρεια των εικόνων είναι κρίσιμες για την αναγνώριση παθολογιών που αφορούν τους μαλακούς ιστούς, περιλαμβανομένων του εγκεφάλου, των μυών και των συνδετικών ιστών.

Η πολυπλοκότητα της ερμηνείας των εικόνων MRI ενισχύεται από την ανάγκη για εξειδικευμένη γνώση και εκπαίδευση, καθώς και από την απαιτητικότητά της για υψηλή ανάλυση αντίθεσης μεταξύ των δομών. Η εφαρμογή της στη νευρολογία είναι εξαιρετικά σημαντική, διασφαλίζοντας τη λεπτομερή εξέταση των δομών του εγκεφάλου και την ανίχνευση ανωμαλιών όπως όγκοι, σκλήρυνση κατά πλάκας και νευροεκφυλιστικές ασθένειες. Ειδικότερα, η λειτουργική MRI (fMRI) μπορεί να παρέχει πληροφορίες για την εγκεφαλική δραστηριότητα, βοηθώντας στη διάγνωση και στη μελέτη καταστάσεων όπως η επιληψία και οι εγκεφαλικές κακώσεις.

Τέλος, οι προηγμένες τεχνικές όπως η απορροφησιμετρία ακτίνων Χ διπλής ενέργειας (DEXA) έχουν επιτρέψει τη λεπτομερή αξιολόγηση της οστικής πυκνότητας, κρίσιμης σημασίας για τη διάγνωση οστεοπόρωσης και άλλων μεταβολικών ασθενειών των οστών.

- **Υπολογιστική Τομογραφία (CT):** Η αξονική τομογραφία (CT) χρησιμοποιεί την τεχνολογία ακτίνων Χ για την απόκτηση λεπτομερών εικόνων διατομής του ανθρώπινου σώματος. Μέσω της συνδυασμένης απεικόνισης πολλαπλών εικόνων ακτίνων Χ από διαφορετικές γωνίες, η αξονική τομογραφία παρέχει λεπτομερείς αναπαραστάσεις των οστών, των αιμοφόρων αγγείων και των μαλακών ιστών. Η σημαντική αυτή τεχνολογία αποτελεί κρίσιμο εργαλείο για τη διάγνωση ποικίλων παθήσεων, συμπεριλαμβανομένων των καρκίνων, των καρδιαγγειακών παθήσεων και των τραυματικών καταστάσεων [11].

Η ανάλυση αυτών των εικόνων απαιτεί εξειδικευμένη εκπαίδευση και εμπειρία, καθώς και εξοικείωση με τον μεγάλο όγκο δεδομένων που παράγεται και την ανάγκη για εντοπισμό λεπτών διαφορών στην πυκνότητα των ιστών. Στην πρακτική, οι αξονικές τομογραφίες χρησιμοποιούνται επίσης για την ανίχνευση καρκίνων του πνεύμονα, αποκαλύπτοντας μικρούς όγκους που θα μπορούσαν να παραμείνουν αόρατοι με μια συνήθη ακτινογραφία θώρακα. Επιπλέον, η αξονική αγγειογραφία είναι κρίσιμη για την αξιολόγηση της ροής του αίματος στις αρτηρίες και τις φλέβες σε όλο το σώμα, συμβάλλοντας στη διάγνωση και στη θεραπεία των αγγειακών παθήσεων.

Με την εισαγωγή των αξονικών τομογραφιών χαμηλής δόσης (LDCT), έχει επιτευχθεί σημαντική πρόοδος στον έλεγχο του καρκίνου του πνεύμονα, επιτρέποντας την έγκαιρη ανίχνευση με μειωμένη έκθεση στην ακτινοβολία.

Η ανάλυση και η ερμηνεία των ψηφιακών ιατρικών εικόνων αποτελούν σημαντική πρόκληση λόγω της μεγάλης ποσότητας δεδομένων και της πολυπλοκότητας των εικόνων που παράγονται. Οι παραδοσιακές προσεγγίσεις εστιάζουν στην εμπειρογνωμοσύνη των ακτινολόγων και των κλινικών ιατρών, οι οποίοι υποβάλλονται σε εκτεταμένη εκπαίδευση για την ανάλυση ιατρικών εικόνων και τον εντοπισμό ανωμαλιών. Η επεξεργασία των εικόνων απαιτεί την αντιμετώπιση προκλήσεων όπως η κίνηση του ασθενούς, η ποιότητα της εικόνας και η παρουσία τεχνουργημάτων, τα οποία μπορούν να επηρεάσουν την ερμηνεία.

Για παράδειγμα, μεταλλικά εμφυτεύματα μπορούν να δημιουργήσουν τεχνουργήματα στις εικόνες αξονικής και μαγνητικής τομογραφίας, προκαλώντας επισκιάσεις και πιθανές παρερμηνείες στην ερμηνεία των λεπτομερειών. Οι πρόσφατες εξελίξεις στη μηχανική μάθηση και την επεξεργασία εικόνας αποσκοπούν στη μείωση της επίδρασης αυτών των τεχνουργημάτων και στη βελτίωση της σαφήνειας των εικόνων.

Η ανάπτυξη προηγμένων αλγορίθμων αναγνώρισης και εξαγωγής χαρακτηριστικών μέσω της τεχνητής νοημοσύνης ανοίγει νέους ορίζοντες για την αυτοματοποίηση και τη βελτίωση της ακρίβειας στην ερμηνεία ιατρικών εικόνων, βοηθώντας έτσι στην ακόμα πιο αποτελεσματική διάγνωση και θεραπεία των ασθενών.

Τα τελευταία χρόνια, η ανάπτυξη της τεχνητής νοημοσύνης (AI) και της βαθιάς μάθησης έχει ωθήσει την καινοτομία στην ανάλυση ιατρικών εικόνων, προσφέροντας τη δυνατότητα ενίσχυσης των ανθρώπινων ικανοτήτων και βελτίωσης των κλινικών ροών εργασίας. Τα μοντέλα βαθιάς μάθησης, ιδίως τα συνελκτικά νευρωνικά δίκτυα (CNN), έχουν επιδείξει αξιοσημείωτη ακρίβεια στον εντοπισμό και την ταξινόμηση αντικειμένων εντός ιατρικών εικόνων [1, 17].

- **AI στην ταξινόμηση εικόνων:** Η ταξινόμηση εικόνων στον τομέα της ιατρικής απεικόνισης αναφέρεται στην επισήμανση εικόνων με τη σωστή περιγραφή του περιεχομένου τους. Αυτό μπορεί να περιλαμβάνει την αναγνώριση παθολογικών καταστάσεων, όπως πνευμονία ή όγκοι στον εγκέφαλο, μέσω της ανάλυσης ακτινογραφιών ή μαγνητικών τομογραφιών.

Η χρήση μοντέλων βαθιάς μάθησης, όπως τα Convolutional Neural Networks (CNNs), έχει αποδειχθεί ότι παρέχει υψηλές επιδόσεις σε αυτές τις εργασίες. Αυτά τα μοντέλα εκπαιδεύονται σε μεγάλα σύνολα δεδομένων με ετικέτες ιατρικής απεικόνισης, προκειμένου να μάθουν να αναγνωρίζουν αυτόματα παθολογίες στις εικόνες με υψηλή ακρίβεια. Η

έρευνα έχει δείξει ότι αυτά τα μοντέλα μπορούν να ξεπεράσουν ακόμα και τους έμπειρους ακτινολόγους στην ακρίβεια της διάγνωσης, όπως στην περίπτωση της πνευμονίας και της φυματίωσης.

Ειδικά σε περιβάλλοντα με περιορισμένους ανθρώπινους πόρους, όπου οι εκπαιδευμένοι ακτινολόγοι είναι σπάνιοι, η τεχνητή νοημοσύνη μπορεί να αποτελέσει πολύτιμο εργαλείο για την προσήλωση και την ακριβή διάγνωση. Η μάθηση μεταφοράς είναι επίσης σημαντική, καθώς επιτρέπει στα μοντέλα να ωφεληθούν από προ-εκπαιδευμένη γνώση, βελτιώνοντας την απόδοση και στις περιπτώσεις όπου τα επισημασμένα δεδομένα είναι περιορισμένα.

Αυτές οι προσεγγίσεις συνδυάζουν την τεχνολογία και την ιατρική γνώση για να προωθήσουν τη διάγνωση και τη θεραπεία, προσφέροντας νέες δυνατότητες για την αντιμετώπιση ιατρικών προκλήσεων.

- **ΑΙ στην ανίχνευση αντικειμένων:** Η ανίχνευση αντικειμένων αναφέρεται στη διαδικασία εντοπισμού και αναγνώρισης πολλαπλών αντικειμένων σε μια εικόνα. Στην ιατρική απεικόνιση, αυτή η δυνατότητα έχει εφαρμοστεί επιτυχώς για την εντοπισμό διαφόρων παθολογικών καταστάσεων, όπως όγκοι, θρόμβοι αίματος και άλλες ανωμαλίες σε ιατρικές εικόνες, όπως αξονικές τομογραφίες και μαγνητικές τομογραφίες.

Αλγόριθμοι ανίχνευσης αντικειμένων όπως τα YOLO (You Only Look Once) και Faster R-CNN έχουν εφαρμοστεί με επιτυχία στην ιατρική απεικόνιση. Αυτά τα μοντέλα μπορούν να αναγνωρίσουν και να επισημάνουν περιοχές ενδιαφέροντος στις εικόνες, παρέχοντας οριοθετημένα πλαίσια γύρω από τα αντικείμενα που ανιχνεύονται. Αυτό βοηθάει τους ακτινολόγους να εστιάσουν στις κρίσιμες περιοχές και να βελτιώσουν τη διαγνωστική ακρίβεια και αποτελεσματικότητα. Οι πληροφορίες που προσφέρουν αυτά τα μοντέλα μπορούν να είναι ιδιαίτερα χρήσιμες για την πρόωγη ανίχνευση και την περαιτέρω εξέταση πιθανών παθολογιών.

Επιπλέον, η συνδυασμένη χρήση τεχνητής νοημοσύνης με την ακτινομετρία (που αναφέρεται στην εξαγωγή ποσοτικών χαρακτηριστικών από ιατρικές εικόνες) έχει βελτιώσει σημαντικά την ικανότητα πρόβλεψης της πρόγνωσης της νόσου και της ανταπόκρισης στη θεραπεία. Αυτή η ενσωμάτωση επιτρέπει την αξιοποίηση της πλήρους πληροφορίας που παρέχουν οι ιατρικές εικόνες, με στόχο την πιο εξειδικευμένη και αποτελεσματική ιατρική περίθαλψη.

- **Το ΑΙ στην τμηματοποίηση:** Η τμηματοποίηση εικόνας στην ιατρική απεικόνιση

αποτελεί έναν σημαντικό τομέα εφαρμογής της τεχνητής νοημοσύνης, καθώς επιτρέπει την ανάλυση και την ακριβή κατάτμηση πολύπλοκων ιατρικών εικόνων σε διάφορα τμήματα με νόημα. Αυτό είναι κρίσιμο για την ακριβή διάγνωση και την εξατομικευμένη θεραπεία.

Μοντέλα τμηματοποίησης βασισμένα σε τεχνητή νοημοσύνη, όπως αυτά που βασίζονται στην αρχιτεκτονική U-Net, έχουν αποδείξει την ικανότητά τους να αναγνωρίζουν και να κατανοούν αυτόματα δομές ενδιαφέροντος σε ιατρικές εικόνες. Αυτά τα μοντέλα μπορούν να τμηματοποιήσουν διαφορετικούς τύπους ιστών, οργάνων και παθολογικές περιοχές, όπως όγκους, οργανικά συστήματα και αιμοφόρα αγγεία. Αυτό επιτρέπει την ακριβή αναγνώριση και ανάλυση περιοχών ενδιαφέροντος, βοηθώντας στην αντιμετώπιση κρίσιμων ιατρικών καταστάσεων.

Επιπλέον, η χρήση τεχνικών μάθησης με ημι-επίβλεψη και χωρίς επίβλεψη στην τμηματοποίηση έχει βελτιώσει τη δυνατότητα των μοντέλων να μάθουν από περιορισμένα επισημασμένα δεδομένα. Αυτό επιτρέπει στα μοντέλα να εξάγουν σημαντικά χαρακτηριστικά και να παράγουν ακριβείς τμηματοποιήσεις ακόμα και όταν τα δεδομένα εκπαίδευσης είναι περιορισμένα. Συνολικά, η τμηματοποίηση εικόνας με τη χρήση της τεχνητής νοημοσύνης ενισχύει την ικανότητα των ιατρών να διαγνώσουν και να θεραπεύσουν ασθένειες με μεγαλύτερη ακρίβεια και προσαρμοστικότητα.

- **Υποτιτλοποίηση ιατρικών εικόνων:** Η υποτιτλοποίηση ιατρικών εικόνων είναι ένας σύγχρονος τομέας που συνδυάζει την όραση υπολογιστών (CV) και την επεξεργασία φυσικής γλώσσας (NLP) με σκοπό την αυτοματοποίηση της δημιουργίας περιγραφικών και κλινικών συνοδευτικών κειμένων για ιατρικές εικόνες. Αυτός ο τομέας έχει ως στόχο την ενίσχυση της ερμηνευσιμότητας των ιατρικών εικόνων και την υποστήριξη της λήψης κλινικών αποφάσεων. Τα μοντέλα υποτιτλισμού ιατρικών εικόνων συνήθως χρησιμοποιούν έναν συνδυασμό των Convolutional Neural Networks (CNN) για την εξαγωγή χαρακτηριστικών από τις εικόνες και Recurrent Neural Networks (RNN), ενισχυμένα με μηχανισμούς προσοχής, για την παραγωγή κειμένου. Αυτά τα μοντέλα προσπαθούν να παράγουν υπότιτλους που όχι μόνο περιγράφουν το οπτικό περιεχόμενο των εικόνων αλλά και παρέχουν σημαντικές πληροφορίες σχετικά με τις κλινικές επιπτώσεις των ευρημάτων.

Η εφαρμογή αυτής της τεχνολογίας μπορεί να είναι εξαιρετικά χρήσιμη σε πολυάσχολες κλινικές συνθήκες, καθώς οι αυτοματοποιημένοι υπότιτλοι μπορούν να παρέχουν προκαταρκτικές αναφορές που επισημαίνουν τα σημαντικά ευρήματα των ιατρικών εικόνων. Αυτό

επιτρέπει στους ακτινολόγους να οργανώνουν καλύτερα το φόρτο εργασίας τους και να επικεντρωθούν σε πιο σύνθετες περιπτώσεις. Επιπλέον, η υποτιτλοποίηση μπορεί να βοηθήσει στην εκπαίδευση νέων ακτινολόγων και στη βελτίωση της εκπαίδευσης στην αναγνώριση και διαχείριση ιατρικών καταστάσεων.

Συνολικά, ο υποτιτλισμός ιατρικών εικόνων αποτελεί μια καινοτόμο εφαρμογή της τεχνητής νοημοσύνης που υπόσχεται να βελτιώσει την ακρίβεια, την αξιοπιστία και την ερμηνευσιμότητα στον τομέα της ιατρικής απεικόνισης.

B.1 Υπότιτλοι εικόνας στην όραση υπολογιστών

B.1.1 Πρώιμες προσεγγίσεις: Μέθοδοι βασισμένες σε πρότυπα

Οι πρώιμες προσεγγίσεις για τη δημιουργία υποτίτλων εικόνων συχνά βασίζονταν σε μεθόδους που χρησιμοποιούσαν προκαθορισμένα πρότυπα για τη γενέτειρα περιγραφή των εικόνων. Αυτές οι μέθοδοι εξάγουν πληροφορίες από την εικόνα με στόχο τη δημιουργία περιγραφών που αντιστοιχούν σε σταθερά πρότυπα. Για παράδειγμα, ένα απλό πρότυπο μπορεί να είναι "Αυτό είναι ένα [αντικείμενο] σε ένα [περιβάλλον]", με τα κενά να συμπληρώνονται από αλγόριθμους αναγνώρισης αντικειμένων [3]. Παρόλο που αυτή η προσέγγιση είναι απλή, περιορίζεται από την έλλειψη ευελιξίας και την αδυναμία να αντιληφθεί τις αποχρώσεις και την πλούσια ποικιλία της φυσικής γλώσσας που χρησιμοποιεί ο άνθρωπος [20].

Συχνά, αυτές οι μέθοδοι παρήγαγαν επαναλαμβανόμενες και τυποποιημένες περιγραφές που ήταν περιορισμένες σε πληροφορίες και δεν είχαν τον πλούτο των περιγραφών που μπορεί να δημιουργήσει ένας ανθρώπινος παρατηρητής [3]. Επιπλέον, οι απλοϊκοί υπότιτλοι που παράγονταν από αυτές τις μεθόδους δυσκολεύονταν να αντιμετωπίσουν πολύπλοκες εικόνες που περιείχαν πολλαπλά αντικείμενα ή σκηνές, με αποτέλεσμα να είναι ελλιπείς ή να μην είναι ικανοποιητικοί για πρακτικές εφαρμογές [20].

Αυτή η προσέγγιση έχει αλλάξει με την πρόοδο της τεχνητής νοημοσύνης, ειδικά με τη χρήση προηγμένων μοντέλων μηχανικής μάθησης που μπορούν να εξάγουν πιο πλούσιες και πιο πληροφοριακές περιγραφές από τις εικόνες, βελτιώνοντας έτσι την εφαρμοστικότητα της υποτιτλοποίησης εικόνων στην πράξη [20]. Ένα από τα πιο σημαντικά βήματα προς αυτή την κατεύθυνση ήταν η ανάπτυξη μοντέλων που συνδυάζουν τα πλεονεκτήματα της όρασης και της φυσικής γλώσσας σε ένα ενιαίο πλαίσιο. Τα μοντέλα αυτά, όπως το Neural Image Caption (NIC), χρησιμοποιούν βαθιά συνελικτικά νευρωνικά δίκτυα (CNN) για την εξαγωγή

χαρακτηριστικών από τις εικόνες και αναδρομικά νευρωνικά δίκτυα (RNN) για τη δημιουργία φυσικών προτάσεων που περιγράφουν τις εικόνες [20]. Το NIC, για παράδειγμα, εκπαιδεύεται να μεγιστοποιεί την πιθανότητα της σωστής περιγραφής δεδομένης της εικόνας, επιτυγχάνοντας αποτελέσματα που πλησιάζουν τις ανθρώπινες επιδόσεις σε διάφορα σύνολα δεδομένων, όπως το Pascal, το Flickr30k και το MSCOCO [20].

Η ενσωμάτωση αυτών των προηγμένων τεχνικών έχει οδηγήσει σε σημαντική βελτίωση της ποιότητας και της ποικιλίας των περιγραφών εικόνων, κάνοντας τις περιγραφές πιο φυσικές και κατανοητές για τους χρήστες. Η χρήση μεθόδων μηχανικής μάθησης, όπως τα LSTM, επιτρέπει στο σύστημα να παρακολουθεί τα αντικείμενα και τις δράσεις που έχουν ήδη περιγραφεί, αποφεύγοντας επαναλήψεις και ενισχύοντας τη συνοχή των περιγραφών [20]. Αυτές οι εξελίξεις ανοίγουν νέες δυνατότητες για την πρακτική εφαρμογή της υποτιτλοποίησης εικόνων, όπως η βοήθεια σε άτομα με προβλήματα όρασης και η βελτίωση της αλληλεπίδρασης ανθρώπου-υπολογιστή [3, 20].

B.1.2 Εξελίξεις στη βαθιά μάθηση

Η ενσωμάτωση των CNN και των RNN σήμανε σημαντική πρόοδο στη λειτουργία υποτίτλων εικόνων, επιτρέποντας στα μοντέλα να δημιουργούν υποτίτλους που είναι όχι μόνο περιγραφικοί αλλά και συμφωνικοί με τα πλαίσια. Τα CNN εξάγουν ιεραρχικά χαρακτηριστικά από τις εικόνες, ανιχνεύοντας χωρικές σχέσεις και πρότυπα σε διαφορετικές κλίμακες. Αυτά τα χαρακτηριστικά αποτελούν πλούσιες εισόδους για τα RNN, τα οποία είναι ικανά στην επεξεργασία ακολουθιών δεδομένων και στη δημιουργία συνεκτικών κειμένων περιγραφής.

Επιπλέον, η εξέλιξη των μηχανισμών προσοχής εντός αυτών των αρχιτεκτονικών ενίσχυσε ακόμα περισσότερο την ποιότητα των δημιουργούμενων υποτίτλων. Οι μηχανισμοί προσοχής επιτρέπουν στα μοντέλα να επικεντρώνονται επιλεκτικά σε σημαντικά τμήματα της εικόνας κατά τη δημιουργία κάθε λέξης του υποτίτλου, βελτιώνοντας έτσι τη συνολική συνοχή και πληροφοριακή αξία των περιγραφών. Αυτή η ικανότητα είναι ιδιαίτερα χρήσιμη σε πολύπλοκες σκηνές όπου υπάρχουν πολλά αντικείμενα ή περιοχές ενδιαφέροντος.

Η συνεργία μεταξύ CNN και RNN, ενισχυμένη από μηχανισμούς προσοχής, έχει επιτρέψει στα πιο προηγμένα συστήματα υποτιτλοποίησης εικόνας να παράγουν υποτίτλους που όχι μόνο περιγράφουν ακριβώς το οπτικό περιεχόμενο, αλλά παρέχουν επίσης εκφραστικές ερμηνείες που συμβαδίζουν στενά με την ανθρώπινη αντίληψη και κατανόηση.

Συμβολικά Νευρωνικά Δίκτυα (CNNs):

Τα CNN έχουν εδραιωθεί ως ο ακρογωνιαίος λίθος της σύγχρονης όρασης υπολογιστών, λόγω της ικανότητάς τους να μαθαίνουν ιεραρχικές αναπαραστάσεις οπτικών δεδομένων. Ένα συνηθισμένο CNN αποτελείται από πολλαπλά επίπεδα συνελίξεων, συγκέντρωσης και μη γραμμικών ενεργοποιήσεων, τα οποία σταδιακά εξάγουν χαρακτηριστικά υψηλότερου επιπέδου από την εικόνα εισόδου. Στον τομέα της δημιουργίας υποτίτλων, τα CNN χρησιμοποιούνται για την ακριβή εξαγωγή σημαντικών οπτικών χαρακτηριστικών, τα οποία αποτελούν τη βάση για την παραγωγή περιγραφικών λεζάντων [13, 9].

Τα βασικά στοιχεία των CNNs για τη δημιουργία λεζάντας εικόνας περιλαμβάνουν:

- **Στρώματα συνέλιξης:** Τα στρώματα συνέλιξης εφαρμόζουν διάφορες πράξεις στην εικόνα εισόδου, όπως πράξεις συνέλιξης και συγκέντρωσης, για την ανίχνευση τοπικών χαρακτηριστικών όπως ακμές, υφές και μοτίβα. Η βαθύτητα των επιπέδων συνέλιξης επιτρέπει στο δίκτυο να μαθαίνει αναπαραστάσεις που γίνονται όλο και πιο αφηρημένες καθώς προχωρά στην ανάλυση της εικόνας. Τα φίλτρα που χρησιμοποιούνται στα στρώματα συνελικτικής ανάλυσης επιτρέπουν την εξαγωγή και αναγνώριση διαφόρων χαρακτηριστικών εντός της εικόνας, επιτρέποντας έτσι στο μοντέλο να κατανοεί το οπτικό περιεχόμενο σε πολλαπλά επίπεδα αφαίρεσης και να δημιουργεί πιο ακριβείς αναπαραστάσεις της εικόνας.
- **Στρώματα συγκέντρωσης:** Τα στρώματα συγκέντρωσης αποτελούν σημαντικό στάδιο στη διαδικασία επεξεργασίας των χαρτών χαρακτηριστικών σε ένα CNN. Οι συγκεντρωτικές τεχνικές, όπως η μέγιστη και η μέση συγκέντρωση, βοηθούν στη μείωση των χωρικών διαστάσεων των χαρτών χαρακτηριστικών, ενώ διατηρούν τα πιο σημαντικά χαρακτηριστικά. Αυτό μειώνει την υπολογιστική πολυπλοκότητα και βελτιώνει την απόδοση του μοντέλου. Η μέγιστη συγκέντρωση επιλέγει το μέγιστο στοιχείο από κάθε περιοχή του χάρτη χαρακτηριστικών, ενώ η μέση συγκέντρωση υπολογίζει τον μέσο όρο των στοιχείων εντός μιας περιοχής. Και οι δύο τεχνικές συμβάλλουν στη σύνοψη των σημαντικών πληροφοριών, κάνοντας το μοντέλο πιο αποτελεσματικό στην αντιμετώπιση ποικίλων εικόνων εισόδου.
- **Πλήρως συνδεδεμένα στρώματα:** Τα υψηλού επιπέδου χαρακτηριστικά που εξάγονται από τα στρώματα συνελίξεων και συγκέντρωσης υποβάλλονται στην επεξεργασία των πλήρως συνδεδεμένων στρωμάτων. Αυτά τα στρώματα μετατρέπουν τα χαρακτηριστικά σε μια διανυσματική αναπαράσταση της εικόνας σταθερού μεγέθους, ενσωματώνοντας τα σε μια ολοκληρωμένη αναπαράσταση που μπορεί να χρησιμοποιηθεί

για τη δημιουργία λεζάντων [11]. Η διαδικασία αυτή επιτρέπει στο μοντέλο να αποκτήσει κατανόηση για το περιεχόμενο της εικόνας και να παράγει πλούσιες και περιγραφικές προτάσεις που αντανακλούν το περιεχόμενο και το πλαίσιο της εικόνας.

Νευρωνικά Δίκτυα Επαναλαμβανόμενου Χρόνου (RNN):

Τα RNN (Ανατροφοδοτούμενα Νευρωνικά Δίκτυα) είναι μια ειδική κατηγορία νευρωνικών δικτύων που είναι κατάλληλα για διαδοχική επεξεργασία δεδομένων, γεγονός που τα καθιστά ιδανικά για τη δημιουργία προτάσεων ή λεζάντων. Σε αντίθεση με τα παραδοσιακά νευρωνικά δίκτυα τροφοδότησης, τα RNN έχουν συνδέσεις που σχηματίζουν κατευθυνόμενους κύκλους, επιτρέποντάς τους να διατηρούν μια κρυφή κατάσταση που συλλαμβάνει πληροφορίες σχετικά με τις προηγούμενες εισόδους. Αυτή η ικανότητα επιτρέπει στα RNN να παράγουν συνεκτικές και σχετικές με το πλαίσιο ακολουθίες κειμένου [21].

Τα RNN έχουν βρει εφαρμογές σε διάφορους τομείς, όπως η επεξεργασία φυσικής γλώσσας, η μετάφραση μεταξύ γλωσσών, η αναγνώριση ομιλίας και η παραγωγή κειμένου. Χάρη στην ικανότητά τους να μαθαίνουν και να μοντελοποιούν ακολουθίες δεδομένων, τα RNN μπορούν να χειριστούν δεδομένα μεταβλητού μήκους, όπως προτάσεις ή παραγράφους κειμένου.

Ωστόσο, τα παραδοσιακά RNN αντιμετωπίζουν προβλήματα όπως το πρόβλημα της εξαφάνισης του βαθμού και της έκρηξης του βαθμού, τα οποία μπορεί να επηρεάσουν την απόδοσή τους κατά τη μοντελοποίηση μακρών ακολουθιών. Για να αντιμετωπιστούν αυτά τα προβλήματα, έχουν προταθεί διάφορες αρχιτεκτονικές, όπως τα LSTM (Long Short-Term Memory) και τα GRU (Gated Recurrent Units), τα οποία έχουν τη δυνατότητα να μαθαίνουν και να διατηρούν μακροπρόθεσμες εξαρτήσεις με πιο αποτελεσματικό τρόπο.

Για τη δημιουργία υποτίτλων σε εικόνες, τα RNN χρησιμοποιούνται για τη δημιουργία της ακολουθίας των λέξεων που συνθέτουν τη λεζάντα. Τα βασικά χαρακτηριστικά περιλαμβάνουν:

- **Embedding Layer:** Το στρώμα ενσωμάτωσης (embedding layer) είναι ένα σημαντικό συστατικό στα μοντέλα RNN για τη δημιουργία λεζάντων εικόνων. Αυτό το στρώμα μετατρέπει κάθε λέξη του λεξιλογίου σε μια πυκνή διανυσματική αναπαράσταση, αποτυπώνοντας το σημασιολογικό νόημα των λέξεων. Οι ενσωματώσεις επιτρέπουν στο μοντέλο να εργάζεται με τις λέξεις ως συνεχή διανύσματα και όχι ως διακριτές μάρκες, επιτρέποντάς του να καταγράφει τις σχέσεις μεταξύ των λέξεων. Οι ενσωματώσεις λέξεων μπορούν είτε να εκπαιδευτούν από την αρχή ως μέρος του μον-

τέλου RNN, είτε να προ-εκπαιδευτούν χρησιμοποιώντας τεχνικές όπως η Word2Vec ή η GloVe σε μεγάλα σώματα κειμένων. Η χρήση προ-εκπαιδευμένων ενσωματώσεων μπορεί να βελτιώσει την απόδοση και να επιταχύνει την εκπαίδευση, καθώς ενσωματώνουν πλούσιες σημασιολογικές πληροφορίες από μεγάλα σύνολα δεδομένων. Επιπλέον, οι ενσωματώσεις μπορούν να είναι στατικές ή να εκπαιδεύονται περαιτέρω κατά τη διάρκεια της εκπαίδευσης του μοντέλου RNN. Τα στατικά ενσωματώματα διατηρούν τις αρχικές τους τιμές, ενώ τα εκπαιδύσιμα ενσωματώματα επιτρέπουν στο μοντέλο να προσαρμόσει τις αναπαραστάσεις λέξεων για τη συγκεκριμένη εργασία. Συνολικά, το στρώμα ενσωμάτωσης παρέχει μια αποτελεσματική αναπαράσταση των λέξεων που επιτρέπει στο RNN να εκμεταλλευτεί τις σημασιολογικές σχέσεις μεταξύ των λέξεων και να παράγει πιο ακριβείς και συναφείς λεζάντες εικόνων.

- **LSTM/GRU Units:** Τα LSTM και GRU είναι δημοφιλείς αρχιτεκτονικές RNN για εργασίες όπως η δημιουργία λεζάντων εικόνων, χάρη στις ικανότητές τους να μοντελοποιούν μακροπρόθεσμες εξαρτήσεις και να αντιμετωπίζουν προβλήματα όπως η εξαφάνιση του βαθμού. Τα δίκτυα μακράς βραχυχρόνιας μνήμης (LSTM) και οι μονάδες ανατροφοδοτούμενων πυλών (GRU) ενσωματώνουν ειδικές πύλες (gates) που ελέγχουν τη ροή της πληροφορίας μέσα στο δίκτυο. Αυτές οι πύλες καθορίζουν ποιες πληροφορίες θα διατηρηθούν στην κρυφή κατάσταση (hidden state) του δικτύου και ποιες θα απορριφθούν ή αποθηκευτούν στη μνήμη του κελιού. Η ικανότητα των LSTM και GRU να ρυθμίζουν αυτές τις πύλες τους επιτρέπει να συλλαμβάνουν εξαρτήσεις μεγάλου εύρους στα δεδομένα, κάτι που είναι ιδιαίτερα σημαντικό για τη δημιουργία συνεκτικών και σχετικών λεζάντων εικόνων. Μπορούν να διατηρούν σημαντικές πληροφορίες από προηγούμενες λέξεις ή εικόνες και να τις χρησιμοποιούν για να προβλέπουν τις επόμενες λέξεις. Επιπλέον, η δυνατότητα να "ξεχνούν" άσχετες πληροφορίες βοηθά στην αντιμετώπιση του προβλήματος της εξαφάνισης του βαθμού, που μπορεί να προκύψει σε παραδοσιακά RNN κατά τη μοντελοποίηση μακρών ακολουθιών. Συνολικά, οι LSTM και GRU μονάδες παρέχουν ισχυρούς μηχανισμούς μνήμης που βελτιώνουν σημαντικά την απόδοση των RNN σε εργασίες όπως η δημιουργία λεζάντων εικόνων, επιτρέποντας στα μοντέλα να παράγουν πιο ακριβείς, συναφείς και συνεκτικές περιγραφές.
- **Attention Mechanisms:** Οι μηχανισμοί προσοχής (attention mechanisms) έχουν συμβάλει σημαντικά στη βελτίωση της απόδοσης των RNN για τη δημιουργία λεζάντων εικόνων. Αυτοί οι μηχανισμοί επιτρέπουν στο μοντέλο να εστιάζει σε διαφορετικές περιοχές της εικόνας σε κάθε χρονικό βήμα κατά τη διαδικασία παραγωγής της λεζάντας.

Πιο συγκεκριμένα, οι μηχανισμοί προσοχής υπολογίζουν ένα διάνυσμα βαρών προσοχής που καθορίζει τη σχετική σημασία κάθε περιοχής της εικόνας για την πρόβλεψη της επόμενης λέξης της λεζάντας. Αυτά τα βάρη προσοχής προσαρμόζονται δυναμικά σε κάθε χρονικό βήμα, επιτρέποντας στο μοντέλο να επικεντρώνεται στα πιο σχετικά χαρακτηριστικά της εικόνας για την τρέχουσα λέξη που πρόκειται να παραχθεί. Η επιλεκτική εστίαση που παρέχουν οι μηχανισμοί προσοχής συμβάλλει στη δημιουργία πιο ακριβών και συναφών με το πλαίσιο λεζάντων. Αντί να βασίζεται σε μια στατική, ολιστική αναπαράσταση της εικόνας, το μοντέλο μπορεί να εστιάσει σε συγκεκριμένα τμήματα που είναι πιο σχετικά με την τρέχουσα λέξη-στόχο. Υπάρχουν διάφοροι τύποι μηχανισμών προσοχής, όπως η ακριβής προσοχή (hard attention) και η μαλακή προσοχή (soft attention), καθώς και πιο σύνθετες παραλλαγές όπως η πολλαπλή προσοχή (multi-head attention). Αυτές οι παραλλαγές διαφέρουν ως προς τον τρόπο με τον οποίο υπολογίζουν και εφαρμόζουν τα βάρη προσοχής. Συνολικά, η ενσωμάτωση μηχανισμών προσοχής στα RNN έχει βελτιώσει σημαντικά την ποιότητα των παραγόμενων λεζάντων εικόνων, επιτρέποντας στα μοντέλα να εστιάζουν στα πιο κρίσιμα χαρακτηριστικά της εικόνας και να παράγουν περιγραφές που είναι πιο ακριβείς, συναφείς και πλούσιες σε λεπτομέρειες.

B.1.3 Μετρικές αξιολόγησης

BLEU (Bilingual Evaluation Understudy): Η BLEU μετρά την επικάλυψη μεταξύ των n -γραμμμάτων στις παραγόμενες λεζάντες και στις λεζάντες αναφοράς. Χρησιμοποιείται ευρέως στη μηχανική μετάφραση και στη δημιουργία λεζάντας εικόνων για την αξιολόγηση της ακρίβειας του παραγόμενου κειμένου. Οι βαθμολογίες BLEU υπολογίζονται συγκρίνοντας την παραγόμενη λεζάντα με μία ή περισσότερες λεζάντες αναφοράς και αξιολογώντας την επικάλυψη σύντομων ακολουθιών λέξεων (n -grams). Οι υψηλότερες βαθμολογίες BLEU υποδηλώνουν καλύτερη ευθυγράμμιση με τις λεζάντες αναφοράς, γεγονός που υποδηλώνει ακριβέστερες και συναφείς περιγραφές [17].

B.1.4 Προκλήσεις και μελλοντικές κατευθύνσεις

Παρά τις προόδους, εξακολουθούν να υπάρχουν αρκετές προκλήσεις στον τομέα του υποτίτλισμού εικόνων, ιδίως σε ιατρικές εφαρμογές:

- **Διαθεσιμότητα δεδομένων:** Τα σύνολα δεδομένων υψηλής ποιότητας με σχολιασμό είναι απαραίτητα για την εκπαίδευση αποτελεσματικών μοντέλων. Στον τομέα του υποτιτλισμού ιατρικών εικόνων, η απόκτηση μεγάλων και διαφορετικών συνόλων δεδομένων με ακριβείς λεζάντες μπορεί να αποτελέσει πρόκληση. Η διαδικασία του χειροκίνητου σχολιασμού ιατρικών εικόνων με λεπτομερείς και κλινικά σχετικές λεζάντες είναι χρονοβόρα και απαιτεί γνώσεις εμπειρογνομόνων. Οι προσπάθειες για τη δημιουργία και την κοινή χρήση τυποποιημένων συνόλων δεδομένων, όπως το σύνολο δεδομένων του Indiana University Chest X-Ray, το σύνολο δεδομένων DDSM για εικόνες μαστογραφίας και το σύνολο δεδομένων ISIC για εικόνες δερματικών βλαβών, παίζουν σημαντικό ρόλο για την πρόοδο του τομέα [5].
- **Εξειδικευμένη γλώσσα:** Ο υποτιτλισμός ιατρικών εικόνων απαιτεί τη χρήση εξειδικευμένης ιατρικής ορολογίας, η οποία απαιτεί γνώσεις και εμπειρογνομosύνη. Τα μοντέλα πρέπει να εκπαιδεύονται ώστε να κατανοούν και να παράγουν κείμενο που αντικατοπτρίζει με ακρίβεια τις ιατρικές έννοιες και τα ευρήματα. Αυτό περιλαμβάνει την αναγνώριση συγκεκριμένων ιατρικών καταστάσεων, ανατομικών όρων και διαδικαστικών λεπτομερειών, καθώς και τη δημιουργία περιγραφών που έχουν νόημα και είναι χρήσιμες σε ένα κλινικό πλαίσιο [1].
- **Ερμηνευσιμότητα:** Η διασφάλιση ότι οι παραγόμενες λεζάντες είναι ερμηνεύσιμες και κλινικά σημαντικές είναι κρίσιμη για την αποδοχή και τη χρήση τους σε περιβάλλοντα υγειονομικής περίθαλψης. Οι κλινικοί γιατροί πρέπει να εμπιστεύονται ότι οι λεζάντες που παράγονται από την TN αντικατοπτρίζουν με ακρίβεια το οπτικό περιεχόμενο των ιατρικών εικόνων και παρέχουν σχετικές πληροφορίες για τη διάγνωση και τη θεραπεία. Οι τεχνικές για τη βελτίωση της ερμηνευσιμότητας, όπως η οπτικοποίηση χαρτών προσοχής και η παροχή εξηγήσεων για τις αποφάσεις του μοντέλου, είναι απαραίτητες για την οικοδόμηση εμπιστοσύνης και αποδοχής μεταξύ των επαγγελματιών υγείας [24].

B.2 CNN στην ιατρική απεικόνιση

Τα Convolutional Neural Networks (CNNs) είναι ιδιαίτερα ικανά στην καταγραφή τόσο των λεπτομερειών χαμηλού επιπέδου όσο και της σημασιολογίας υψηλού επιπέδου, γεγονός που τα καθιστά ιδανικά για την ανάλυση σύνθετων ιατρικών εικόνων. Η ιεραρχική φύση των CNNs, που χαρακτηρίζεται από πολλαπλά στρώματα convolutional φίλτρων, pooling operations και

μη γραμμικές ενεργοποιήσεις, τους επιτρέπει να δημιουργούν προοδευτικά πιο αφηρημένες αναπαραστάσεις των οπτικών δεδομένων, από ακμές και υφές έως πλήρεις ανατομικές δομές και παθολογικά μοτίβα.

Οι πρόσφατες εξελίξεις στις αρχιτεκτονικές των CNNs, όπως οι ResNet, DenseNet και EfficientNet, έχουν βελτιώσει περαιτέρω την απόδοσή τους, αντιμετωπίζοντας ζητήματα όπως το vanishing gradients και βελτιστοποιώντας την υπολογιστική αποδοτικότητα. Αυτές οι αρχιτεκτονικές ενσωματώνουν καινοτόμες τεχνικές όπως residual connections, dense connections και compound scaling, αντίστοιχα, οι οποίες βοηθούν στην καλύτερη εξαγωγή χαρακτηριστικών και στην ακρίβεια του μοντέλου.

Ο συνδυασμός των CNNs με άλλες τεχνικές μηχανικής μάθησης, όπως το reinforcement learning και τα generative adversarial networks (GANs), επίσης δείχνει υποσχόμενα αποτελέσματα. Για παράδειγμα, τα GANs μπορούν να δημιουργήσουν συνθετικές ιατρικές εικόνες για την αύξηση των συνόλων εκπαίδευσης, βοηθώντας στην υπέρβαση του περιορισμού των ανεπαρκών δεδομένων με ετικέτες στην ιατρική απεικόνιση.

- **Εξόρυξη χαρακτηριστικών χαμηλού επιπέδου:** Στα αρχικά επίπεδα, τα CNN επικεντρώνονται στην εξαγωγή χαρακτηριστικών χαμηλού επιπέδου, όπως ακμές, γωνίες και υφές. Τα χαρακτηριστικά αυτά συλλαμβάνονται μέσω πράξεων συνέλιξης που εφαρμόζουν διάφορα φίλτρα στην εικόνα εισόδου. Κάθε φίλτρο ανιχνεύει συγκεκριμένα μοτίβα σε μικρές τοπικές περιοχές της εικόνας, επιτρέποντας στο δίκτυο να εντοπίζει θεμελιώδη οπτικά στοιχεία [13, 9]. Καθώς προχωράμε σε βαθύτερα επίπεδα του δικτύου, τα CNN αρχίζουν να αναγνωρίζουν πιο σύνθετα και αφηρημένα χαρακτηριστικά. Σε αυτά τα ενδιαμέσια επίπεδα, το δίκτυο συνδυάζει τα χαμηλού επιπέδου χαρακτηριστικά για να ανιχνεύσει συνδυασμούς μοτίβων, όπως σχήματα και απλές ανατομικές δομές. Στα τελικά επίπεδα, τα CNN είναι σε θέση να εξάγουν υψηλού επιπέδου χαρακτηριστικά που αντιπροσωπεύουν πλήρως τις σύνθετες ανατομικές δομές και τα παθολογικά μοτίβα. Τα χαρακτηριστικά αυτά είναι πιο αφηρημένα και περιλαμβάνουν σημαντικές πληροφορίες για τη διάγνωση και την κατανόηση των ιατρικών εικόνων. Επιπλέον, τα CNN χρησιμοποιούν τεχνικές όπως pooling operations, οι οποίες μειώνουν τη διάσταση των δεδομένων, διατηρώντας παράλληλα τις σημαντικές πληροφορίες. Το max pooling, για παράδειγμα, επιλέγει το μέγιστο στοιχείο από μια περιοχή χαρακτηριστικών, βοηθώντας στην αντοχή του δικτύου σε μικρές παραλλαγές και μετατοπίσεις στην εικόνα. Η χρήση των μη γραμμικών ενεργοποιήσεων, όπως οι ReLU (Rectified Linear

Units), επιτρέπει στα CNN να μοντελοποιούν σύνθετες μη γραμμικές σχέσεις μεταξύ των χαρακτηριστικών. Αυτό βελτιώνει την ικανότητα του δικτύου να μαθαίνει από τα δεδομένα και να γενικεύει σε νέα, άορατα δείγματα. Τα CNN έχουν αποδειχθεί ιδιαίτερα αποτελεσματικά σε διάφορες εφαρμογές ιατρικής απεικόνισης, όπως η ανίχνευση όγκων σε ακτινογραφίες και μαγνητικές τομογραφίες, η ταξινόμηση δερματικών βλαβών και η ανάλυση παθολογικών εικόνων. Η συνεχής έρευνα και οι τεχνολογικές εξελίξεις αναμένεται να βελτιώσουν περαιτέρω την απόδοση και την εφαρμοσιμότητα των CNNs στην ιατρική διάγνωση και θεραπεία.

- **Convolutional layers:** Αυτά τα στρώματα εφαρμόζουν convolutional φίλτρα στην εικόνα εισόδου, παράγοντας feature maps που αναδεικνύουν σημαντικές οπτικές ενδείξεις. Τα πρώιμα convolutional layers συλλαμβάνουν βασικά χαρακτηριστικά όπως οι ακμές και οι υφές, τα οποία είναι απαραίτητα για τη διάκριση μεταξύ διαφορετικών τύπων ιστών και δομών σε ιατρικές εικόνες. Τα φίλτρα σε αυτά τα στρώματα σαρώνουν την εικόνα με την τακτική του sliding window, ανιχνεύοντας μοτίβα και χαρακτηριστικά ζωτικής σημασίας για την αρχική κατανόηση της εικόνας.
- **Activation functions:** Μη γραμμικές activation functions, όπως η ReLU (Rectified Linear Unit), εφαρμόζονται στους feature maps για την εισαγωγή μη γραμμικότητας, επιτρέποντας στο δίκτυο να μοντελοποιεί σύνθετα μοτίβα και σχέσεις στα δεδομένα. Η ενεργοποίηση ReLU συμβάλλει στον μετριασμό του προβλήματος της εξαφανιζόμενης κλίσης, επιτρέποντας στο δίκτυο να μαθαίνει πιο αποτελεσματικά κατά τη διάρκεια της εκπαίδευσης [13, 21].

Εξόρυξη χαρακτηριστικών υψηλού επιπέδου

Καθώς τα δεδομένα προχωρούν μέσα από τα βαθύτερα στρώματα του CNN, το δίκτυο αρχίζει να συλλαμβάνει χαρακτηριστικά υψηλού επιπέδου, συμπεριλαμβανομένων πιο σύνθετων μοτίβων και δομών. Αυτά τα χαρακτηριστικά υψηλότερου επιπέδου είναι κρίσιμα για τον εντοπισμό ανατομικών δομών, την ανίχνευση ανωμαλιών και τη διάγνωση ασθενειών [13, 7].

- **Pooling Layers:** Τα pooling layers, όπως η max pooling ή η average pooling, χρησιμοποιούνται για τη μείωση των χωρικών διαστάσεων των feature maps, διατηρώντας παράλληλα τις πιο σημαντικές πληροφορίες. Η διαδικασία αυτή βοηθάει στο να γίνει το δίκτυο πιο αποδοτικό από υπολογιστική άποψη και πιο ανθεκτικό στις μεταβολές

της εικόνας εισόδου. Τα pooling layers συγκεντρώνουν τις πληροφορίες από μικρές περιοχές των feature maps, διασφαλίζοντας ότι διατηρούνται τα πιο σημαντικά χαρακτηριστικά.

- **Fully Connected Layers:** Τα υψηλού επιπέδου χαρακτηριστικά που εξάγονται από τα convolutional και pooling layers τελικά ισοπεδώνονται και τροφοδοτούνται σε fully connected layers. Αυτά τα στρώματα ενσωματώνουν τα εξαγόμενα χαρακτηριστικά και εκτελούν τις τελικές εργασίες ταξινόμησης ή παλινδρόμησης, όπως ο εντοπισμός της παρουσίας ενός όγκου ή η κατάτμηση ενός οργάνου. Τα fully connected layers συνδυάζουν τα αφηρημένα χαρακτηριστικά σε μια ολοκληρωμένη αναπαράσταση, επιτρέποντας ακριβείς προβλέψεις [13, 5].

Εφαρμογές των CNN στην ιατρική απεικόνιση

Τα CNN έχουν εφαρμοστεί σε ένα ευρύ φάσμα εργασιών ιατρικής απεικόνισης, επιδεικνύοντας αξιοσημείωτες επιδόσεις και ακρίβεια σε διάφορες εφαρμογές:

- **Image Classification:** Τα CNN μπορούν να ταξινομήσουν ιατρικές εικόνες σε διάφορες κατηγορίες, όπως φυσιολογικές ή μη φυσιολογικές, καλοήθειες ή κακοήθειες. Για παράδειγμα, ένα CNN μπορεί να εκπαιδευτεί για να διακρίνει μεταξύ υγιούς πνευμονικού ιστού και πνευμονίας σε ακτινογραφίες θώρακος. Αυτή η ικανότητα ταξινόμησης είναι σημαντική για την έγκαιρη ανίχνευση ασθενειών και την άμεση παρέμβαση [13, 7, 5].
- **Object Detection:** Τα CNN μπορούν να ανιχνεύσουν και να εντοπίσουν συγκεκριμένα αντικείμενα σε μια εικόνα, όπως όγκους, κατάγματα ή βλάβες. Αυτή η ικανότητα είναι ιδιαίτερα χρήσιμη στη μαστογραφία, όπου τα CNN μπορούν να επισημάνουν ύποπτες περιοχές που μπορεί να υποδηλώνουν την παρουσία καρκίνου του μαστού. Μοντέλα ανίχνευσης αντικειμένων, όπως το Faster R-CNN και το YOLO, χρησιμοποιούνται για τον εντοπισμό και την επισήμανση περιοχών ενδιαφέροντος εντός ιατρικών εικόνων, διευκολύνοντας την εστιασμένη ανάλυση από τους ακτινολόγους [1, 8].
- **Image Segmentation:** Τα CNN μπορούν να τμηματοποιήσουν ιατρικές εικόνες για να οριοθετήσουν διαφορετικές ανατομικές δομές ή παθολογικές περιοχές. Για παράδειγμα, στην ιστοπαθολογία, τα CNN μπορούν να διαχωρίσουν τα καρκινικά κύτταρα από τα υγιή κύτταρα, παρέχοντας πολύτιμες πληροφορίες για τη διάγνωση και το σχεδιασμό της

θεραπείας. Μοντέλα τμηματοποίησης, όπως το U-Net και οι παραλλαγές του, χρησιμοποιούνται για την παραγωγή σχολίων ανά εικονοστοιχείο που είναι απαραίτητα για την ακριβή και λεπτομερή ιατρική ανάλυση [15, 26].

Προκλήσεις και καινοτομίες

Παρά την επιτυχία τους, τα CNN στην ιατρική απεικόνιση αντιμετωπίζουν αρκετές προκλήσεις. Μία από τις πρωταρχικές προκλήσεις είναι η περιορισμένη διαθεσιμότητα σχολιασμένων συνόλων ιατρικών δεδομένων. Τα σύνολα δεδομένων υψηλής ποιότητας με ετικέτες είναι ζωτικής σημασίας για την εκπαίδευση ακριβών μοντέλων CNN, αλλά η απόκτηση τέτοιων συνόλων δεδομένων μπορεί να είναι χρονοβόρα και δαπανηρή [4, 19]. Επιπλέον, οι ιατρικές εικόνες παρουσιάζουν συχνά μεγάλη μεταβλητότητα λόγω των διαφορών στα πρωτόκολλα απεικόνισης, τις συνθήκες των ασθενών και τον εξοπλισμό, γεγονός που μπορεί να επηρεάσει την απόδοση των μοντέλων CNN [13, 26].

Μια άλλη πρόκληση είναι η ανάγκη για ερμηνευσιμότητα και διαφάνεια στα αποτελέσματα των CNN. Οι κλινικοί γιατροί πρέπει να κατανοούν και να εμπιστεύονται τις προβλέψεις που κάνουν τα μοντέλα για να τα χρησιμοποιήσουν στην κλινική πράξη. Η χρήση τεχνικών όπως οι heatmaps και οι attention mechanisms μπορεί να βοηθήσει στην κατανόηση των αποφάσεων του μοντέλου, αλλά αυτές οι μέθοδοι χρειάζονται περαιτέρω ανάπτυξη και αποδοχή στην ιατρική κοινότητα.

Επιπλέον, τα CNN συχνά απαιτούν μεγάλη υπολογιστική ισχύ για την εκπαίδευση και την ανάλυση των δεδομένων. Αυτό μπορεί να είναι πρόβλημα για μικρότερα ιατρικά κέντρα ή για περιοχές με περιορισμένους πόρους. Η ανάπτυξη αποδοτικών αλγορίθμων και η χρήση εξειδικευμένου υλικού, όπως οι GPUs, μπορούν να βοηθήσουν στην αντιμετώπιση αυτής της πρόκλησης.

Για την αντιμετώπιση αυτών των προκλήσεων, οι ερευνητές έχουν διερευνήσει διάφορες στρατηγικές, συμπεριλαμβανομένης της αύξησης των δεδομένων, της μάθησης μεταφοράς και της χρήσης μηχανισμών προσοχής:

- **Data Augmentation:** Τεχνικές όπως η περιστροφή (rotation), η κλιμάκωση (scaling) και η αναστροφή (flipping) χρησιμοποιούνται για την τεχνητή αύξηση του μεγέθους και της ποικιλομορφίας του συνόλου δεδομένων εκπαίδευσης. Αυτό βοηθά στη βελτίωση της ευρωστίας και της γενικευσιμότητας των μοντέλων CNN. Η data augmentation

μιμείται τις φυσικές παραλλαγές που απαντώνται στα δεδομένα του πραγματικού κόσμου, επιτρέποντας στο μοντέλο να μαθαίνει αμετάβλητα χαρακτηριστικά και να αποδίδει καλά σε αθέατες εικόνες [1, 4].

- **Transfer Learning:** Τα προ-εκπαιδευμένα μοντέλα CNN, όπως αυτά που εκπαιδεύονται σε μεγάλα σύνολα γενικών δεδομένων εικόνων όπως το ImageNet, μπορούν να ρυθμιστούν λεπτομερώς σε σύνολα δεδομένων ιατρικών εικόνων. Αυτή η προσέγγιση αξιοποιεί τα μαθημένα χαρακτηριστικά από το προ-εκπαιδευμένο μοντέλο, μειώνοντας την ανάγκη για μεγάλα σχολιασμένα σύνολα ιατρικών δεδομένων. Η transfer learning έχει αποδειχθεί ιδιαίτερα αποτελεσματική σε σενάρια όπου τα ιατρικά δεδομένα με ετικέτες είναι σπάνια, επιτρέποντας στα μοντέλα να επωφεληθούν από τις πλούσιες αναπαραστάσεις χαρακτηριστικών που μαθαίνονται από μη ιατρικές εικόνες [14, 19].
- **Attention Mechanisms:** Οι attention mechanisms επιτρέπουν στα CNN να εστιάζουν στις πιο σχετικές περιοχές της εικόνας, ενισχύοντας την ικανότητά τους να συλλαμβάνουν κρίσιμες λεπτομέρειες. Αυτό είναι ιδιαίτερα χρήσιμο στην ιατρική απεικόνιση, όπου συγκεκριμένες περιοχές ενδιαφέροντος, όπως όγκοι ή βλάβες, πρέπει να εντοπιστούν και να αναλυθούν με ακρίβεια. Οι attention mechanisms ρυθμίζουν δυναμικά την εστίαση του μοντέλου, εξασφαλίζοντας ότι δίνει μεγαλύτερη προσοχή σε διαγνωστικά σημαντικές περιοχές κατά τη δημιουργία προβλέψεων [6, 25].

Η ενσωμάτωση αυτών των προηγμένων τεχνικών έχει βελτιώσει σημαντικά την απόδοση και τη δυνατότητα εφαρμογής των CNN στην ιατρική απεικόνιση, ανοίγοντας το δρόμο για πιο ακριβή και αποτελεσματικά διαγνωστικά εργαλεία. Οι καινοτομίες αυτές όχι μόνο ενισχύουν την ικανότητα του μοντέλου να γενικεύει σε διαφορετικά σύνολα δεδομένων και κλινικά σενάρια, αλλά συμβάλλουν επίσης στην ανάπτυξη πιο αξιόπιστων και ερμηνεύσιμων συστημάτων AI για εφαρμογές υγειονομικής περίθαλψης.

RNN για την παραγωγή κειμένου

Τα επαναλαμβανόμενα νευρωνικά δίκτυα (Recurrent Neural Networks - RNNs) είναι αναπόσπαστο μέρος της παραγωγής κειμένου της λεζάντας ιατρικών εικόνων. Τα RNN, ιδίως εκείνα που είναι εξοπλισμένα με Long Short-Term Memory (LSTM) ή Gated Recurrent Units (GRUs), είναι ικανά να συλλάβουν τις διαδοχικές εξαρτήσεις εντός των δεδομένων εισόδου, καθιστώντας τα κατάλληλα για τη δημιουργία συνεκτικών και συναφών με το περιεχόμενο

κειμενικών περιγραφών. Αυτά τα δίκτυα έχουν σχεδιαστεί για να επεξεργάζονται ακολουθίες δεδομένων διατηρώντας μια hidden state που καταγράφει πληροφορίες σχετικά με τις προηγούμενες εισόδους, η οποία είναι απαραίτητη για εργασίες που περιλαμβάνουν language modeling και text generation.

B.2.1 Sequential Data Processing

Τα Recurrent Neural Networks (RNNs) επεξεργάζονται δεδομένα εισόδου με διαδοχικό τρόπο, όπου κάθε στοιχείο εισόδου εξαρτάται από τα προηγούμενα. Αυτό επιτρέπει στα RNNs να διατηρούν μια memory προηγούμενων εισόδων, η οποία είναι ζωτικής σημασίας για τη δημιουργία συνεκτικού και σχετικού με τα συμφραζόμενα κειμένου [16, 15].

- **Hidden State:** Σε κάθε χρονικό βήμα, το RNN ενημερώνει την hidden state του με βάση την τρέχουσα είσοδο και την προηγούμενη hidden state. Αυτή η hidden state λειτουργεί ως memory, η οποία καταγράφει πληροφορίες σχετικά με την ακολουθία που έχει δει μέχρι στιγμής. Η hidden state ενημερώνεται συνεχώς ώστε να αντικατοπτρίζει το πλαίσιο ολόκληρης της ακολουθίας, επιτρέποντας στο RNN να παράγει κείμενο που διατηρεί τη συνάφεια και τη συνοχή του πλαισίου [13, 11].
- **LSTM/GRU Units:** Οι μονάδες LSTM και GRU είναι εξειδικευμένες αρχιτεκτονικές RNN που έχουν σχεδιαστεί για την αντιμετώπιση του προβλήματος της εξαφανιζόμενης κλίσης, το οποίο μπορεί να εμποδίσει την εκμάθηση εξαρτήσεων μεγάλης εμβέλειας. Αυτές οι μονάδες διαθέτουν μηχανισμούς πύλης που ελέγχουν τη ροή των πληροφοριών, επιτρέποντας στο δίκτυο να διατηρεί σημαντικές πληροφορίες για μεγάλες ακολουθίες. Οι πύλες στις μονάδες LSTM και GRU (input gate, forget gate και output gate) βοηθούν στην επιλεκτική ενημέρωση και διατήρηση της πληροφορίας, καθιστώντας τις αποτελεσματικές για εργασίες που απαιτούν μακροχρόνια μνήμη [17, 8].

B.2.2 Δημιουργία κειμένου σε ιατρικές εικόνες με λεζάντες

Στη δημιουργία λεζάντας σε ιατρικές εικόνες, τα RNN χρησιμοποιούνται για τη δημιουργία της ακολουθίας των λέξεων που αποτελούν τη λεζάντα με βάση τα χαρακτηριστικά που εξάγονται από τα CNN. Η διαδικασία περιλαμβάνει:

- **Feature Encoding:** Ο χάρτης χαρακτηριστικών που παράγεται από το CNN χρησι-

μπορείται ως αρχική είσοδος για το RNN. Αυτός ο χάρτης χαρακτηριστικών ενσωματώνει τις οπτικές πληροφορίες που είναι απαραίτητες για τη δημιουργία της λεζάντας. Τα κωδικοποιημένα χαρακτηριστικά παρέχουν μια ολοκληρωμένη αναπαράσταση της εικόνας, καταγράφοντας τόσο τις λεπτομέρειες χαμηλού όσο και υψηλού επιπέδου που είναι απαραίτητες για την ακριβή περιγραφή [17, 8].

- **Word Embeddings:** Κάθε λέξη του λεξιλογίου αναπαρίσταται ως πυκνό διάνυσμα σε έναν χώρο υψηλών διαστάσεων, αποτυπώνοντας τη σημασιολογική της σημασία. Αυτές οι ενσωματώσεις λέξεων χρησιμοποιούνται ως είσοδοι για το RNN κατά τη διαδικασία παραγωγής λεζάντας. Οι ενσωματώσεις εξασφαλίζουν ότι οι σημασιολογικά παρόμοιες λέξεις τοποθετούνται κοντά η μία στην άλλη στο διανυσματικό χώρο, βοηθώντας στη δημιουργία ουσιαστικών και κατάλληλων από άποψη πλαισίου προτάσεων [15, 16].
- **Sequential Word Generation:** Το RNN παράγει τη λεζάντα λέξη προς λέξη, ενημερώνοντας την hidden state του σε κάθε χρονικό βήμα. Η επιλογή κάθε επόμενης λέξης επηρεάζεται από τις προηγούμενες λέξεις και τα κωδικοποιημένα οπτικά χαρακτηριστικά, διασφαλίζοντας τη συνοχή και τη συνάφεια με το πλαίσιο. Η διαδοχική φύση των RNNs τους επιτρέπει να δημιουργούν προτάσεις που αντικατοπτρίζουν το οπτικό περιεχόμενο με ακρίβεια και ευχέρεια [11, 13].

B.2.3 Μηχανισμοί προσοχής

Οι μηχανισμοί προσοχής έχουν βελτιώσει σημαντικά την απόδοση των RNNs σε εργασίες παραγωγής κειμένου. Στο πλαίσιο της δημιουργίας λεζάντας εικόνας, οι μηχανισμοί προσοχής επιτρέπουν στο μοντέλο να εστιάζει σε διαφορετικά μέρη της εικόνας σε κάθε χρονικό βήμα, βελτιώνοντας την ακρίβεια και τη συνάφεια των παραγόμενων λεζάντων [17].

- **Selective Attention:** Σε κάθε χρονικό βήμα, ο μηχανισμός attention υπολογίζει ένα σύνολο βαρών που υποδηλώνουν τη σημασία διαφορετικών περιοχών της εικόνας. Αυτά τα βάρη χρησιμοποιούνται για τη δημιουργία ενός σταθμισμένου αθροίσματος του χάρτη χαρακτηριστικών, το οποίο στη συνέχεια χρησιμοποιείται ως είσοδος για το RNN. Αυτό επιτρέπει στο μοντέλο να επικεντρώνεται στα πιο σημαντικά τμήματα της εικόνας κατά τη δημιουργία κάθε λέξης της λεζάντας [15, 16].
- **Dynamic Attention:** Τα βάρη attention ενημερώνονται δυναμικά σε κάθε χρονικό βήμα, επιτρέποντας στο μοντέλο να προσαρμόζει την εστίασή του με βάση το πλαίσιο

της μέχρι τώρα παραγόμενης λεζάντας. Αυτή η δυναμική εστίαση βοηθά στη σύλληψη των πιο σχετικών οπτικών λεπτομερειών για κάθε μέρος της λεζάντας, διασφαλίζοντας ότι το παραγόμενο κείμενο αντικατοπτρίζει με ακρίβεια τα βασικά χαρακτηριστικά της εικόνας [11].

B.2.4 Εκπαίδευση RNN για τη λεζάντα ιατρικών εικόνων

Η εκπαίδευση RNN για τον υποτιτλισμό ιατρικών εικόνων περιλαμβάνει διάφορα κρίσιμα βήματα:

- **Data Pairs:** Η εκπαίδευση απαιτεί μεγάλα σύνολα δεδομένων ιατρικών εικόνων σε συνδυασμό με ακριβείς λεζάντες. Αυτά τα σύνολα δεδομένων χρησιμοποιούνται για να διδάξουν στο μοντέλο τις συσχετίσεις μεταξύ οπτικών χαρακτηριστικών και γλωσσικών περιγραφών. Τα σύνολα δεδομένων υψηλής ποιότητας είναι απαραίτητα για την καταγραφή της μεταβλητότητας της ιατρικής απεικόνισης και τη διασφάλιση ότι οι παραγόμενες λεζάντες είναι κλινικά σχετικές [11].
- **Supervised Learning:** Το RNN εκπαιδεύεται χρησιμοποιώντας μάθηση με επίβλεψη, όπου το μοντέλο μαθαίνει να προβλέπει την επόμενη λέξη της λεζάντας δεδομένων των προηγούμενων λέξεων και των κωδικοποιημένων χαρακτηριστικών της εικόνας. Η διαδικασία εκπαίδευσης περιλαμβάνει την ελαχιστοποίηση μιας συνάρτησης απωλειών, όπως η cross-entropy, η οποία μετρά τη διαφορά μεταξύ των προβλεπόμενων και των πραγματικών λέξεων. Η supervised learning εξασφαλίζει ότι το μοντέλο μαθαίνει από δεδομένα με ακριβή σχολιασμό, βελτιώνοντας την απόδοσή του σε αθέατες εικόνες [8].
- **Evaluation Metrics:** Η ποιότητα των παραγόμενων λεζάντων αξιολογείται με τη χρήση μετρικών όπως οι BLEU, METEOR και CIDEr, οι οποίες συγκρίνουν τις παραγόμενες λεζάντες με τις λεζάντες αναφοράς για να αξιολογήσουν την ακρίβεια και τη συνάφεια τους. Αυτές οι μετρικές παρέχουν ποσοτικά μέτρα της απόδοσης του μοντέλου, αναδεικνύοντας τα δυνατά του σημεία και τους τομείς που χρήζουν βελτίωσης.

B.2.5 Προκλήσεις και μελλοντικές κατευθύνσεις

Παρόλο που τα RNN έχουν αποδειχθεί αποτελεσματικά για τη δημιουργία υποτίτλων ιατρικών εικόνων, παραμένουν αρκετές προκλήσεις:

- **Ελλείψεις δεδομένων:** Τα υψηλής ποιότητας ζεύγη δεδομένων είναι απαραίτητα για την εκπαίδευση, αλλά συχνά είναι σπάνια στον ιατρικό τομέα. Η ανάπτυξη μεθόδων για την επαύξηση ή τη συνθετική παραγωγή δεδομένων εκπαίδευσης θα μπορούσε να βοηθήσει στην αντιμετώπιση αυτού του ζητήματος. Η έλλειψη δεδομένων μπορεί να περιορίσει την ικανότητα του μοντέλου να γενικεύει σε διαφορετικές ιατρικές καταστάσεις και τρόπους απεικόνισης [11].
- **Γλώσσα εξειδικευμένη στο πεδίο:** Η ιατρική γλώσσα είναι εξαιρετικά εξειδικευμένη και εξαρτάται από το πλαίσιο. Η διασφάλιση ότι τα RNN μπορούν να παράγουν με ακρίβεια κλινικά σχετικές και ακριβείς περιγραφές απαιτεί εκτεταμένη γνώση του τομέα. Η ενσωμάτωση οντολογιών και ορολογιών ειδικών για τον τομέα στη διαδικασία εκπαίδευσης μπορεί να βελτιώσει την ακρίβεια και τη συνάφεια των παραγόμενων λεζάντων [15].
- **Interpretability:** Η φύση του μαύρου κουτιού των RNN μπορεί να καταστήσει δύσκολη την κατανόηση του τρόπου λήψης συγκεκριμένων αποφάσεων. Η ενίσχυση της interpretability αυτών των μοντέλων είναι ζωτικής σημασίας για την απόκτηση εμπιστοσύνης σε κλινικά περιβάλλοντα. Τεχνικές όπως η οπτικοποίηση της attention και τα model explanation frameworks μπορούν να βοηθήσουν τους κλινικούς ιατρούς να κατανοήσουν τη συλλογιστική πίσω από τις παραγόμενες λεζάντες, ενισχύοντας την εμπιστοσύνη και την αποδοχή.

B.3 Μηχανισμοί προσοχής

Οι μηχανισμοί προσοχής έχουν γίνει θεμελιώδη στοιχεία στις σύγχρονες αρχιτεκτονικές βαθιάς μάθησης, ιδίως σε εργασίες που περιλαμβάνουν διαδοχικά ή δομημένα δεδομένα. Αυτοί οι μηχανισμοί αναπτύχθηκαν αρχικά για τη μηχανική μετάφραση, αλλά έκτοτε έχουν εφαρμοστεί σε ένα ευρύ φάσμα εργασιών, συμπεριλαμβανομένων των λεζάντων εικόνων, της αναγνώρισης ομιλίας και άλλων. Η βασική ιδέα πίσω από τους μηχανισμούς προσοχής είναι να επιτρέπουν στο μοντέλο να εστιάζει στα πιο σημαντικά μέρη της ακολουθίας εισόδου κατά τη δημιουργία μιας εξόδου. Αυτή η επιλεκτική εστίαση επιτρέπει στο μοντέλο να κατανέμει διαφορετικά βάρη σε διαφορετικά μέρη της ακολουθίας, ενισχύοντας έτσι την ικανότητά του να παράγει πιο ακριβείς και συναφείς με το πλαίσιο εξόδους [15].

B.3.1 Προέλευση και ανάπτυξη

Η έννοια των μηχανισμών προσοχής εισήχθη για πρώτη φορά στο πλαίσιο της μηχανικής μετάφρασης. Τα παραδοσιακά μοντέλα ακολουθίας-σε-ακολουθία, τα οποία βασίζονταν αποκλειστικά σε επαναλαμβανόμενα νευρωνικά δίκτυα (RNN), αντιμετώπιζαν δυσκολίες στη σύλληψη εξαρτήσεων μεγάλης εμβέλειας και στον χειρισμό μεγάλων ακολουθιών εισόδου. Ο μηχανισμός προσοχής αντιμετώπισε αυτές τις προκλήσεις επιτρέποντας στο μοντέλο να εστιάζει δυναμικά σε συγκεκριμένα τμήματα της ακολουθίας εισόδου, βελτιώνοντας την ποιότητα και την απόδοση της μετάφρασης [17].

Η βασική ιδέα είναι ο υπολογισμός ενός σταθμισμένου αθροίσματος των χαρακτηριστικών εισόδου, όπου τα βάρη (σκορ προσοχής) καθορίζονται με βάση τη συνάφεια κάθε χαρακτηριστικού εισόδου με το τρέχον βήμα εξόδου. Αυτό επιτρέπει στο μοντέλο να επικεντρώνεται στα πιο σημαντικά μέρη της εισόδου κατά την πραγματοποίηση προβλέψεων [15, 16].

Οι μηχανισμοί προσοχής έχουν επίσης εφαρμοστεί ευρέως σε άλλους τομείς πέρα από τη μηχανική μετάφραση, όπως η επεξεργασία φυσικής γλώσσας (NLP), η αναγνώριση ομιλίας, και η ιατρική απεικόνιση. Στην επεξεργασία φυσικής γλώσσας, οι μηχανισμοί προσοχής χρησιμοποιούνται για την ανάλυση κειμένων, την εξαγωγή πληροφορίας και τη σύνοψη κειμένων. Στην αναγνώριση ομιλίας, βοηθούν το μοντέλο να επικεντρώνεται σε κρίσιμα τμήματα της ηχητικής ακολουθίας, βελτιώνοντας την ακρίβεια της αναγνώρισης.

B.3.2 Τύποι μηχανισμών προσοχής

- **Additive (Bahdanau) Attention:** Εισήχθη από τους Bahdanau κ.ά. (2015), η additive attention υπολογίζει τα σκορ προσοχής χρησιμοποιώντας ένα νευρωνικό δίκτυο τροφοδότησης με ένα μόνο κρυφό στρώμα. Αυτός ο μηχανισμός υπολογίζει τις βαθμολογίες ευθυγράμμισης λαμβάνοντας υπόψη τις κρυφές καταστάσεις του κωδικοποιητή και του αποκωδικοποιητή και στη συνέχεια εφαρμόζει μια συνάρτηση softmax για να λάβει τα βάρη προσοχής [12].
- **Multiplicative (Dot-Product) Attention:** Προτεινόμενη από τους Luong et al. (2015), η multiplicative attention, επίσης γνωστή ως dot-product attention, υπολογίζει τα σκορ προσοχής λαμβάνοντας το γινόμενο τελείας των κρυφών καταστάσεων του κωδικοποιητή και της κρυφής κατάστασης του αποκωδικοποιητή. Αυτή η μέθοδος είναι υπολογιστικά αποδοτική και λειτουργεί καλά για πολλές εργασίες [12].

- **Scaled Dot-Product Attention:** Χρησιμοποιείται στο μοντέλο Transformer από τους Vaswani et al. (2017), η scaled dot-product attention τροποποιεί την dot-product attention κλιμακώνοντας τις βαθμολογίες προσοχής με το αντίστροφο της τετραγωνικής ρίζας της διάστασης των διανυσμάτων κλειδιών. Αυτό βοηθά στη σταθεροποίηση των κλίσεων κατά τη διάρκεια της εκπαίδευσης, ειδικά όταν η διαστατικότητα είναι υψηλή.
- **Self-Attention:** Η self-attention, ή ενδοπροσοχή, επιτρέπει στο μοντέλο να συσχετίζει διαφορετικές θέσεις μιας ενιαίας ακολουθίας προκειμένου να υπολογίσει μια αναπαράσταση της ακολουθίας. Αυτός ο μηχανισμός χρησιμοποιείται εκτενώς στην αρχιτεκτονική Transformer, όπου επιτρέπει στο μοντέλο να συλλάβει εξαρτήσεις μεταξύ όλων των στοιχείων της ακολουθίας εισόδου ταυτόχρονα [15].

B.3.3 Εφαρμογή στον υποτιτλισμό ιατρικών εικόνων

Στο πλαίσιο της λεζάντας ιατρικών εικόνων, οι μηχανισμοί προσοχής διαδραματίζουν κρίσιμο ρόλο στην ενίσχυση της ικανότητας του μοντέλου να παράγει λεζάντες που είναι όχι μόνο ακριβείς αλλά και διαγνωστικά σημαντικές. Οι ιατρικές εικόνες συχνά περιέχουν πολλαπλές περιοχές ενδιαφέροντος, καθεμία από τις οποίες έχει διαφορετικό βαθμό κλινικής σημασίας. Οι μηχανισμοί προσοχής επιτρέπουν στο μοντέλο να εστιάζει σε αυτές τις κρίσιμες περιοχές, διασφαλίζοντας ότι οι παραγόμενες λεζάντες είναι κλινικά σχετικές και κατατοπιστικές [15].

- **Visual Attention:** Στον υποτιτλισμό ιατρικής εικόνας, οι μηχανισμοί visual attention επιτρέπουν στο μοντέλο να εστιάζει επιλεκτικά σε διαφορετικά μέρη της εικόνας κατά τη δημιουργία κάθε λέξης στον υποτιτλισμό. Αυτό είναι ιδιαίτερα σημαντικό στην ιατρική απεικόνιση, όπου ορισμένες περιοχές της εικόνας μπορεί να περιέχουν κρίσιμες διαγνωστικές πληροφορίες. Για παράδειγμα, σε μια ακτινογραφία θώρακα, ο μηχανισμός προσοχής μπορεί να βοηθήσει το μοντέλο να εστιάσει σε περιοχές που παρουσιάζουν πιθανές ανωμαλίες, όπως οζίδια ή βλάβες [15, 16].
- **Hierarchical Attention:** Ορισμένα προηγμένα μοντέλα χρησιμοποιούν μηχανισμούς hierarchical attention, όπου η προσοχή εφαρμόζεται σε πολλαπλά επίπεδα της αναπαράστασης της εικόνας. Αυτό επιτρέπει στο μοντέλο να εντοπίζει πρώτα ευρείες περιοχές ενδιαφέροντος και στη συνέχεια να εστιάζει σε λεπτότερες λεπτομέρειες εντός αυτών των περιοχών, βελτιώνοντας την ακρίβεια και τη λεπτομέρεια των παραγόμενων λεζάντων [11].

- **Temporal Attention:** Για διαδοχικά δεδομένα, όπως μελέτες ιατρικής απεικόνισης που περιλαμβάνουν δεδομένα χρονοσειρών (π.χ. υπερηχογραφήματα ή μαγνητικές τομογραφίες με την πάροδο του χρόνου), οι μηχανισμοί temporal attention μπορούν να χρησιμοποιηθούν για την εστίαση στα πιο σημαντικά χρονικά σημεία. Αυτό βοηθάει στη δημιουργία λεζάντων που περιγράφουν με ακρίβεια τις αλλαγές με την πάροδο του χρόνου, κάτι που είναι ζωτικής σημασίας για την παρακολούθηση της εξέλιξης της νόσου ή της ανταπόκρισης στη θεραπεία [11].

B.3.4 Οφέλη και αντίκτυπος

Οι μηχανισμοί προσοχής ενισχύουν σημαντικά την απόδοση των μοντέλων υποτιτλισμού ιατρικών εικόνων βελτιώνοντας την ικανότητά τους να παράγουν ακριβείς, συναφείς με το πλαίσιο και κλινικά σημαντικές περιγραφές. Τα βασικά οφέλη περιλαμβάνουν:

- **Improved Accuracy:** Με την εστίαση στα πιο σχετικά μέρη της εικόνας, οι μηχανισμοί attention βοηθούν το μοντέλο να παράγει ακριβέστερες περιγραφές. Αυτό μειώνει την πιθανότητα να λείπουν κρίσιμα ευρήματα και βελτιώνει τη διαγνωστική ακρίβεια [15].
- **Improved Interpretability:** Οι μηχανισμοί attention καθιστούν τη διαδικασία λήψης αποφάσεων του μοντέλου πιο ερμηνεύσιμη, επισημαίνοντας ποια μέρη της εικόνας συνέβαλαν στη δημιουργία συγκεκριμένων λέξεων στον υποτιτλισμό. Αυτή η διαφάνεια είναι ζωτικής σημασίας για την απόκτηση της εμπιστοσύνης των επαγγελματιών υγείας [16].
- **Reduced Cognitive Load:** Η αυτοματοποιημένη παραγωγή περιγραφικών λεζάντων με μηχανισμούς attention μπορεί να μειώσει το γνωστικό φορτίο των ακτινολόγων και των κλινικών ιατρών, επιτρέποντάς τους να επικεντρωθούν σε πιο σύνθετες εργασίες. Αυτό μπορεί να οδηγήσει σε αυξημένη αποδοτικότητα και μειωμένη κόπωση [15].
- **Easier Communication:** Οι ακριβείς και λεπτομερείς λεζάντες που παράγονται από μοντέλα με ενισχυμένη attention μπορούν να βελτιώσουν την επικοινωνία μεταξύ των επαγγελματιών του τομέα της υγειονομικής περίθαλψης, διασφαλίζοντας ότι οι κρίσιμες πληροφορίες μεταφέρονται αποτελεσματικά και με σαφήνεια [11].

B.4 Υβριδικές αρχιτεκτονικές CNN-RNN

Τα υβριδικά μοντέλα CNN-RNN έχουν χρησιμοποιηθεί ευρέως για να συνδυάσουν τη χωρική κατανόηση των Convolutional Neural Networks (CNN) με τις δυνατότητες διαδοχικής μοντελοποίησης των Recurrent Neural Networks (RNN). Αυτά τα μοντέλα έχουν δείξει σημαντικές υποσχέσεις για τη δημιουργία συνεκτικών και σχετικών με το περιεχόμενο λεζάντων για ιατρικές εικόνες.

Τα CNNs είναι ιδιαίτερα αποτελεσματικά στην εξαγωγή χαρακτηριστικών από τις εικόνες λόγω της ικανότητάς τους να συλλαμβάνουν χωρικές ιεραρχίες στα δεδομένα εικόνες μέσω συνελικτικών και pooling στρωμάτων. Αυτά τα χαρακτηριστικά μπορούν να περιλαμβάνουν βασικά στοιχεία όπως ακμές, γωνίες και υφές, καθώς και πιο σύνθετες δομές όπως όργανα και παθολογικά ευρήματα.

B.5 Προηγούμενες μελέτες

Πολυάριθμες μελέτες έχουν διεξαχθεί στον τομέα της λεζάντας ιατρικών εικόνων, χρησιμοποιώντας ποικίλες μεθοδολογίες και αρχιτεκτονικές. Οι ερευνητές έχουν διερευνήσει διάφορα μοντέλα βαθιάς μάθησης, όπως υβριδικές αρχιτεκτονικές CNN-RNN, μοντέλα προσοχής και μετασχηματιστές, για να αντιμετωπίσουν τις μοναδικές προκλήσεις της λεζάντας ιατρικών εικόνων. Οι μελέτες αυτές συνέβαλαν σημαντικά στην πρόοδο του τομέα, αναδεικνύοντας τις δυνατότητες των προσεγγίσεων που βασίζονται στην τεχνητή νοημοσύνη στη δημιουργία περιγραφικών και κλινικά σχετικών λεζάντων για ιατρικές εικόνες.

- **Neural Captioning for the ImageCLEF 2017 Medical Image Challenges:**

Η μελέτη αυτή κατέδειξε την αποτελεσματικότητα ενός υβριδικού μοντέλου CNN-RNN στο πλαίσιο των προκλήσεων ImageCLEF 2017. Το μοντέλο χρησιμοποίησε CNN για την εξαγωγή χαρακτηριστικών από ιατρικές εικόνες και RNN για τη δημιουργία αντίστοιχων λεζάντων. Τα αποτελέσματα έδειξαν υψηλές βαθμολογίες σε μετρικές αξιολόγησης όπως οι BLEU, METEOR και CIDEr, αναδεικνύοντας την ικανότητα του μοντέλου να παράγει ακριβείς και κατατοπιστικές λεζάντες [13].

- **Efficient Evolving Deep Ensemble Medical Image Captioning Network:**

Η έρευνα αυτή εισήγαγε ένα εξελισσόμενο βαθύ σύνολο μοντέλου που ενσωμάτωσε πολλαπλές αρχιτεκτονικές CNN-RNN. Η προσέγγιση του συνόλου είχε ως στόχο να αξι-

ιοποιήσει τα πλεονεκτήματα των διαφορετικών μοντέλων για να βελτιώσει τη δημιουργία λεζάντας. Η μελέτη ανέφερε βελτιώσεις στην ποιότητα και την ευρωστία των λεζάντων, επιτυγχάνοντας υψηλές βαθμολογίες στις τυπικές μετρικές αξιολόγησης. Η μέθοδος ensemble επέδειξε επίσης καλύτερη γενίκευση σε διάφορες μεθόδους ιατρικής απεικόνισης [21].

B.6 Μοντέλα προσοχής

- Contextualized Keyword Representations for Multi-modal Retinal Image Captioning:** Αυτή η μελέτη διερεύνησε τη χρήση αναπαραστάσεων λέξεων-κλειδιών με βάση το πλαίσιο σε συνδυασμό με μηχανισμούς προσοχής για τη βελτίωση της λεζάντας εικόνων αμφιβληστροειδούς. Η ικανότητα του μοντέλου να εστιάζει σε κρίσιμες περιοχές των εικόνων του αμφιβληστροειδούς οδήγησε σε βελτιωμένη ερμηνευσιμότητα και κλινική συνάφεια των παραγόμενων λεζάντων. Η χρήση μηχανισμών προσοχής επέτρεψε στο μοντέλο να αναδείξει διαγνωστικά σημαντικές περιοχές, παρέχοντας πιο ουσιαστικές περιγραφές [9].
- Expert-defined Keywords Improve Interpretability of Retinal Image Captioning:** Σε αυτή την έρευνα, οι συγγραφείς ενσωμάτωσαν λέξεις-κλειδιά που ορίζονται από εμπειρογνώμονες στον μηχανισμό προσοχής για να βελτιώσουν την ερμηνευσιμότητα των παραγόμενων λεζάντων. Με την ενσωμάτωση της γνώσης του τομέα απευθείας στο μοντέλο, η μελέτη κατέδειξε σημαντικές βελτιώσεις στην ακρίβεια και τη συνάφεια των λεζάντων, καθιστώντας τις πιο χρήσιμες για κλινικούς σκοπούς [7].

B.7 Μετασχηματιστές

Οι μετασχηματιστές έχουν αναδειχθεί ως μια ισχυρή εναλλακτική λύση στα παραδοσιακά RNN, αξιοποιώντας μηχανισμούς αυτοπροσοχής (self-attention) για την καταγραφή των παγκόσμιων εξαρτήσεων στις ακολουθίες δεδομένων. Η ικανότητά τους να επεξεργάζονται παράλληλα ολόκληρες ακολουθίες τους καθιστά ιδιαίτερα αποδοτικούς και αποτελεσματικούς για εργασίες όπως η δημιουργία λεζάντας εικόνων.

Οι μετασχηματιστές, όπως το μοντέλο Transformer που εισήχθη από τους Vaswani et al. (2017), έχουν αλλάξει ριζικά την προσέγγιση της επεξεργασίας φυσικής γλώσσας και άλλων

σχετικών πεδίων. Ο μηχανισμός self-attention επιτρέπει στο μοντέλο να δίνει διαφορετικά βάρη σε κάθε μέρος της εισόδου για κάθε λέξη εξόδου, λαμβάνοντας υπόψη τις εξαρτήσεις μακρινών λέξεων με μεγάλη ακρίβεια. Αυτή η ικανότητα είναι κρίσιμη για την κατανόηση του συμφραζόμενου και τη δημιουργία συνεκτικών κειμένων.

Ένα σημαντικό πλεονέκτημα των μετασχηματιστών είναι η ταχύτητα εκπαίδευσής τους, καθώς η παράλληλη επεξεργασία ακολουθιών μειώνει σημαντικά το χρόνο εκπαίδευσης σε σύγκριση με τα RNNs που επεξεργάζονται δεδομένα διαδοχικά. Αυτή η παράλληλη προσέγγιση επιτρέπει τη χρήση μεγαλύτερων συνόλων δεδομένων και πιο σύνθετων μοντέλων, βελτιώνοντας την απόδοση σε ποικίλες εφαρμογές.

Επιπλέον, οι μετασχηματιστές έχουν αποδειχθεί αποτελεσματικοί σε πολλές εφαρμογές εκτός της επεξεργασίας φυσικής γλώσσας, όπως η δημιουργία λεζάντας για εικόνες (image captioning), η μετάφραση γλώσσας, η αναγνώριση ομιλίας και η ιατρική απεικόνιση. Στην ιατρική απεικόνιση, οι μετασχηματιστές χρησιμοποιούνται για την ανάλυση μεγάλων και σύνθετων συνόλων δεδομένων, επιτρέποντας την ανίχνευση λεπτομερειών που μπορεί να μην είναι άμεσα ορατές με παραδοσιακές μεθόδους.

- **Cross Encoder-Decoder Transformer with Global-Local Visual Extractor for Medical Image Captioning:** Η μελέτη αυτή εισήγαγε ένα μοντέλο μετασχηματιστή διασταυρούμενου κωδικοποιητή-αποκωδικοποιητή που χρησιμοποιεί τόσο παγκόσμια όσο και τοπικά οπτικά χαρακτηριστικά για τη λεζάντα ιατρικής εικόνας. Ο μηχανισμός αυτοπροσοχής της αρχιτεκτονικής του μετασχηματιστή επέτρεψε στο μοντέλο να συλλάβει εξαρτήσεις μεγάλης εμβέλειας και να παράγει λεπτομερείς και ακριβείς απόψεις πλαισίου λεζάντες. Η μελέτη ανέφερε ανώτερες επιδόσεις σε σύγκριση με τα παραδοσιακά μοντέλα που βασίζονται σε RNN, ιδίως στον χειρισμό πολύπλοκων ιατρικών εικόνων [11].
- **A Medical Semantic-Assisted Transformer for Radiographic Report Generation:** Σε αυτή την έρευνα, οι συγγραφείς ανέπτυξαν ένα μοντέλο μετασχηματιστή που ενσωμάτωσε σημασιολογικές πληροφορίες για να βοηθήσει στη δημιουργία ακτινογραφικών εκθέσεων. Με την ενσωμάτωση της ιατρικής σημασιολογίας στην αρχιτεκτονική του μετασχηματιστή, το μοντέλο παρήγαγε λεζάντες που δεν ήταν μόνο ακριβείς αλλά και κλινικά σημαντικές. Η μελέτη ανέδειξε τις δυνατότητες των μετασχηματιστών στην ενίσχυση της ερμηνευσιμότητας και της χρησιμότητας των λεζάντων ιατρικών εικόνων που παράγονται με τεχνητή νοημοσύνη [21].

B.8 Προκλήσεις και μελλοντικές κατευθύνσεις

Παρά την πρόοδο που επιτεύχθηκε σε αυτές τις μελέτες, παραμένουν αρκετές προκλήσεις, οι οποίες πρέπει να αντιμετωπιστούν για να προχωρήσει περαιτέρω ο τομέας του υποτιτλισμού ιατρικών εικόνων:

- **Μεγαλύτερα και πιο ποικίλα σύνολα δεδομένων:** Η ανάπτυξη ισχυρών μοντέλων υποτιτλισμού ιατρικών εικόνων απαιτεί μεγάλα και ποικίλα σύνολα δεδομένων που αποτυπώνουν τη μεταβλητότητα των δεδομένων ιατρικής απεικόνισης. Η απόκτηση τέτοιων συνόλων δεδομένων μπορεί να αποτελέσει πρόκληση λόγω των ανησυχιών για την προστασία της ιδιωτικής ζωής και της ανάγκης για σχολιασμό από εμπειρογνώμονες. Η μελλοντική έρευνα θα πρέπει να επικεντρωθεί στη δημιουργία και την επιμέλεια ολοκληρωμένων συνόλων δεδομένων για τη βελτίωση της γενίκευσης και της απόδοσης των μοντέλων [1, 11].
- **Βελτιωμένες τεχνικές προσαρμογής τομέα:** Τα μοντέλα που εκπαιδεύονται σε συγκεκριμένα σύνολα δεδομένων μπορεί να δυσκολεύονται να γενικευτούν σε διαφορετικές μεθόδους ιατρικής απεικόνισης ή κλινικές ρυθμίσεις. Η ανάπτυξη προηγμένων τεχνικών προσαρμογής σε τομείς μπορεί να βοηθήσει τα μοντέλα να προσαρμοστούν σε νέους τομείς με ελάχιστη επανεκπαίδευση, ενισχύοντας τη δυνατότητα εφαρμογής τους σε ποικίλα κλινικά περιβάλλοντα [24].
- **Improved Interpretability:** Η διασφάλιση ότι οι υπότιτλοι που παράγονται από την ΑΙ είναι ερμηνεύσιμες και κλινικά σημαντικές είναι ζωτικής σημασίας για την αποδοχή τους στην υγειονομική περίθαλψη. Οι μελλοντικές μελέτες θα πρέπει να επικεντρωθούν στη βελτίωση της διαφάνειας των μοντέλων ΑΙ και στην ανάπτυξη μεθόδων για την εξήγηση της λογικής πίσω από τις προβλέψεις τους. Αυτό θα συμβάλει στην οικοδόμηση εμπιστοσύνης μεταξύ των επαγγελματιών του τομέα της υγειονομικής περίθαλψης και θα διευκολύνει την ενσωμάτωση των τεχνολογιών ΑΙ στην κλινική πρακτική [11, 21].
- **Ηθικές και ρυθμιστικές εκτιμήσεις:** Η ανάπτυξη της ΤΝ στην υγειονομική περίθαλψη πρέπει να τηρεί τις ηθικές κατευθυντήριες γραμμές και τα ρυθμιστικά πρότυπα για να διασφαλίζεται η ασφάλεια των ασθενών και η προστασία της ιδιωτικής ζωής των δεδομένων. Οι ερευνητές θα πρέπει να συνεργαστούν στενά με τους ρυθμιστικούς φορείς για την ανάπτυξη πλαισίων που θα αντιμετωπίζουν τις ηθικές ανησυχίες και θα

προωθούν την υπεύθυνη χρήση της ΤΝ στον υποτιτλισμό των ιατρικών εικόνων [11, 21].

Κεφάλαιο C

Μεθοδολογία

C.1 Επισκόπηση

Στην παρούσα εργασία παρουσιάζεται ένα υβριδικό μοντέλο Συνελικτικού Νευρωνικού Δικτύου (CNN) και Επαναλαμβανόμενου Νευρωνικού Δικτύου (RNN) με μηχανισμό προσοχής με σκοπό τον υποτιτλισμό ιατρικών εικόνων. Το στοιχείο CNN είναι υπεύθυνο για την εξαγωγή οπτικών χαρακτηριστικών από τις ιατρικές εικόνες, ενώ το στοιχείο RNN παράγει περιγραφικές και κλινικά σχετικές λεζάντες με βάση αυτά τα χαρακτηριστικά. Αυτό το σύστημα διπλής εισόδου αξιοποιεί τα πλεονεκτήματα τόσο των CNN στην επεξεργασία οπτικών δεδομένων όσο και των RNN στη διαδοχική παραγωγή δεδομένων. Το μοντέλο εκπαιδεύτηκε σε ένα σύνολο δεδομένων 50.000 ιατρικών εικόνων και αξιολογήθηκε χρησιμοποιώντας τη μετρική BLEU. Ο πρωταρχικός στόχος αυτής της μεθοδολογίας είναι η ανάπτυξη ενός στιβαρού και αποδοτικού μοντέλου που να μπορεί να ερμηνεύει σύνθετες ιατρικές εικόνες και να παράγει ακριβείς περιγραφές κειμένου που μπορούν να βοηθήσουν τους επαγγελματίες υγείας στη διάγνωση και τον προγραμματισμό της θεραπείας. Στις ακόλουθες ενότητες παρέχεται λεπτομερής περιγραφή των βημάτων συλλογής και προεπεξεργασίας δεδομένων, του σχεδιασμού του μοντέλου σύντηξης CNN-RNN, της διαδικασίας εκπαίδευσης, των μετρικών αξιολόγησης που χρησιμοποιήθηκαν και της πειραματικής διάταξης.

C.2 Συλλογή δεδομένων

Το σύνολο δεδομένων που χρησιμοποιήθηκε προήλθε από το αποθετήριο MediCat[18],

Περιγραφή	Λεπτομέρειες
Εικόνες από ανοιχτής πρόσβασης άρθρα	217,060 εικόνες από 131,410 ανοιχτής πρόσβασης άρθρα
Σχολιασμοί υπολεζάντων και υποεικόνων	7507 σχολιασμοί υπολεζάντων και υποεικόνων για 2069 εικόνες
Ενσωματωμένες αναφορές	Ενσωματωμένες αναφορές για περίπου 25K εικόνες στο σύνολο δεδομένων ROCO

Πίνακας C.1: Λεπτομέρειες Συνόλου Δεδομένων

Για την παρούσα μελέτη, επιλέχθηκε ένα υποσύνολο 50.000 εικόνων, επιλέγοντας συγκεκριμένα εκείνες με μέγιστο μέγεθος περιγραφής τις 20 λέξεις, ώστε να εξασφαλιστεί η ομοιογένειά του μοντέλου και να παραχθούν πλούσιες και κατατοπιστικές λεζάντες.

C.3 Προεπεξεργασία

C.3.1 Προεπεξεργασία εικόνας

- **Resizing:** Όλες οι εικόνες άλλαξαν μέγεθος σε σταθερή διάσταση (224x224 pixels) για να εξασφαλιστεί ομοιομορφία σε όλο το σύνολο δεδομένων [17].
- **Normalization:** Οι τιμές των εικονοστοιχείων κανονικοποιήθηκαν σε ένα εύρος [0, 1] για τη βελτίωση της σύγκλισης του μοντέλου κατά τη διάρκεια της εκπαίδευσης. Η κανονικοποίηση βοηθά στη μείωση της εσωτερικής μετατόπισης των συνδιακυμάνσεων, οδηγώντας σε ταχύτερη εκπαίδευση και καλύτερη απόδοση [17].

C.3.2 Προεπεξεργασία κειμένου

- **Lowercasing and Punctuation Removal:** Όλες οι λέξεις μετατράπηκαν σε πεζά και αφαιρέθηκαν τα σημεία στίξης και ειδικοί χαρακτήρες για την τυποποίηση των δεδομένων κειμένου. Αυτό βοηθά στη δημιουργία μιας συνεπούς και καθαρής εισόδου για το μοντέλο [21].
- **Tokenization:** Οι λεζάντες υποδιαιρέθηκαν σε μεμονωμένες λέξεις χρησιμοποιώντας έναν tokenizer που χωρίζει το κείμενο με βάση τα κενά [21].

- **Vocabulary Creation:** Δημιουργήθηκε ένα vocabulary από τις λεζάντες εκπαίδευσης, αντιστοιχίζοντας κάθε μοναδική λέξη σε έναν ακέραιο δείκτη. Αυτή η αντιστοίχιση είναι απαραίτητη για τη μετατροπή των λέξεων σε αριθμητικές αναπαραστάσεις που μπορεί να επεξεργαστεί το RNN. Το μέγεθος του vocabulary καθορίστηκε με βάση τη συχνότητα των λέξεων στις λεζάντες, διασφαλίζοντας ότι οι σπάνιες λέξεις είτε συμπεριλαμβάνονταν σε ένα ειδικό token για άγνωστες λέξεις είτε αποκλείονταν εντελώς [21].
- **Sequence Padding:** Οι λεζάντες συμπληρώθηκαν σε σταθερό μήκος για να εξασφαλιστεί η ομοιομορφία στις ακολουθίες εισόδου. Η συμπλήρωση βοηθά στον χειρισμό των διαφορετικών μηκών των λεζάντων, επιτρέποντας στο μοντέλο να τις επεξεργάζεται αποτελεσματικά. Κάθε λεζάντα συμπληρώθηκε με το ειδικό σύμβολο <START> και <END> ώστε να αντιστοιχεί στο μήκος της μεγαλύτερης λεζάντας της παρτίδας [17].

C.4 Αρχιτεκτονική μοντέλου

Η παρεχόμενη αρχιτεκτονική μοντέλου αντιπροσωπεύει μια εξελιγμένη και ισχυρή προσέγγιση στη μάθηση πολλαπλών μοντέλων, ενσωματώνοντας τόσο τα χαρακτηριστικά εικόνas όσο και τα δεδομένα ακολουθίας για ένα ολοκληρωμένο σύστημα πρόβλεψης. Το μοντέλο περιλαμβάνει διακριτά συστατικά για τη διαχείριση χαρακτηριστικών εικόνas και δεδομένων ακολουθίας, καταλήγοντας σε ένα συγχωνευμένο στρώμα αποκωδικοποιητή που εξάγει τις τελικές προβλέψεις.

C.4.1 Encoder - Image Feature Model

Η ενότητα κωδικοποιητή που είναι αφιερωμένη στα χαρακτηριστικά εικόνas ξεκινά με ένα στρώμα Input σχεδιασμένο να αποδέχεται διανύσματα μεγέθους 4096, που μπορεί να αντιπροσωπεύουν προ-εξαγόμενα χαρακτηριστικά από ένα convolutional neural network (CNN) ή έναν άλλο μηχανισμό εξαγωγής χαρακτηριστικών εικόνas. Η πρώτη μεταμόρφωση που εφαρμόζεται σε αυτές τις εισόδους είναι ένα στρώμα Dropout με ρυθμό 0.5, το οποίο χρησιμεύει για την αντιμετώπιση της υπερπροσαρμογής θέτοντας τυχαία το μισό των μονάδων εισόδου σε μηδέν κατά την εκπαίδευση. Μετά το dropout, ένα στρώμα Dense με 512 μονάδες και ReLU ενεργοποίηση συμπιέζει τα χαρακτηριστικά εισόδου σε ένα πιο διαχειρίσιμο μέγεθος, εισάγοντας ταυτόχρονα μη γραμμικότητα στο μοντέλο.

C.4.2 Encoder - Sequence Feature Model

Το μοντέλο χαρακτηριστικών ακολουθίας είναι σχεδιασμένο να διαχειρίζεται δεδομένα ακολουθίας, όπως κείμενο ή χρονοσειρές. Ξεκινά με ένα στρώμα Input που αποδέχεται ακολουθίες καθορισμένου μέγιστου μήκους. Οι ακολουθίες περνούν μέσω ενός στρώματος Embedding, που μετατρέπει τα διακριτά tokens σε πυκνά διανύσματα μεγέθους 256, επιτρέποντας στο μοντέλο να μάθει embeddings λέξεων. Για την καταπολέμηση της υπερπροσαρμογής, ακολουθεί ένα στρώμα Dropout με ρυθμό 0.5. Ο πυρήνας της επεξεργασίας ακολουθιών είναι ένα στρώμα Bidirectional LSTM με 256 μονάδες, που επεξεργάζεται τα δεδομένα ακολουθίας και προς τις δύο κατευθύνσεις, συλλαμβάνοντας έτσι πιο αποτελεσματικά τις συμφραζόμενες πληροφορίες. Η έξοδος αυτού του στρώματος διατηρεί τη μορφή ακολουθίας λόγω της παραμέτρου *return_sequences=True*.

C.4.3 Attention Mechanism

Για την ενίσχυση της εστίασης του μοντέλου σε διαφορετικά μέρη της ακολουθίας, εφαρμόζεται ένας μηχανισμός Attention στην έξοδο του bidirectional LSTM. Ο μηχανισμός attention υπολογίζει ένα context vector που αντιπροσωπεύει τη σημασία κάθε μέρους της ακολουθίας. Ο context vector στη συνέχεια υποβάλλεται σε GlobalAveragePooling1D, το οποίο λαμβάνει το μέσο όρο των δεδομένων ακολουθίας κατά μήκος της χρονικής διάστασης, παράγοντας μια έξοδο σταθερού μεγέθους ανεξάρτητα από το μήκος της εισόδου.

C.4.4 Decoder

Ο αποκωδικοποιητής ενσωματώνει τα επεξεργασμένα χαρακτηριστικά εικόνες και ακολουθίας. Ένα στρώμα Add συνδυάζει τα πυκνά χαρακτηριστικά εικόνες και την ομαδοποιημένη έξοδο attention, εξασφαλίζοντας ότι και οι δύο μορφές δεδομένων συμβάλλουν στην τελική πρόβλεψη. Αυτή η συνδυασμένη αναπαράσταση στη συνέχεια περνά μέσω ενός στρώματος Dense με 256 μονάδες και ReLU ενεργοποίηση, ακολουθούμενο από ένα άλλο στρώμα Dropout με ρυθμό 0.5 για περαιτέρω κανονικοποίηση του μοντέλου. Το τελικό στρώμα είναι ένα στρώμα Dense με softmax ενεργοποίηση, που εξάγει πιθανότητες πάνω από το μέγεθος του λεξιλογίου, κατάλληλο για ταξινομητικά καθήκοντα.

C.4.5 Model Compilation

Το μοντέλο συντίθεται χρησιμοποιώντας τον Adam optimizer με ένα learning rate schedule εκθετικής αποσύνθεσης, ξεκινώντας από $1e-4$ και μειούμενο κατά 0.9 κάθε 10000 βήματα, εξασφαλίζοντας μια πιο συντηρητική και σταθερή διαδικασία εκπαίδευσης. Η συνάρτηση απώλειας που χρησιμοποιείται είναι η categorical cross-entropy, κατάλληλη για πολυταξινόμητικά καθήκοντα. Οι μετρήσεις που παρακολουθούνται κατά την εκπαίδευση περιλαμβάνουν την ακρίβεια, την ακρίβεια (precision) και την ανάκληση (recall), παρέχοντας μια συνολική αξιολόγηση της απόδοσης του μοντέλου.

C.4.6 Callbacks

Πολλές callbacks ενσωματώνονται για να βελτιώσουν την αποτελεσματικότητα της εκπαίδευσης και την απόδοση του μοντέλου:

- Το ReduceLROnPlateau παρακολουθεί την απώλεια επικύρωσης και μειώνει το learning rate κατά συντελεστή 0.2 αν δεν παρατηρηθεί βελτίωση για πέντε epochs, με ένα ελάχιστο όριο learning rate $1e-5$.
- Το ModelCheckpoint αποθηκεύει το μοντέλο μόνο όταν η απώλεια επικύρωσης βελτιώνεται, εξασφαλίζοντας ότι διατηρείται το καλύτερο μοντέλο.
- Το EarlyStopping σταματά την εκπαίδευση αν η απώλεια επικύρωσης δε βελτιωθεί για δέκα συνεχόμενα epochs, αποτρέποντας την υπερπροσαρμογή και την περιττή υπολογιστική διαδικασία.
- Το TensorBoard καταγράφει μετρήσεις εκπαίδευσης, παρέχοντας μια οπτική αναπαράσταση της διαδικασίας εκπαίδευσης.

Αυτή η αρχιτεκτονική μοντέλου επιδεικνύει μια ισχυρή ενσωμάτωση επεξεργασίας δεδομένων εικόνων και ακολουθίας, αξιοποιώντας προηγμένες τεχνικές όπως bidirectional LSTMs, μηχανισμούς attention, και εξελιγμένες στρατηγικές κανονικοποίησης και βελτιστοποίησης για την επίτευξη ισχυρών και ακριβών προβλέψεων.

C.5 CNN για εξαγωγή χαρακτηριστικών

- **Convolutional Layers:** Χρησιμοποιήθηκαν πολλαπλά στρώματα συνελικτικής ανάλυσης για τη σύλληψη μιας σειράς οπτικών χαρακτηριστικών. Τα πρώιμα στρώματα επικεντρώθηκαν σε χαρακτηριστικά χαμηλού επιπέδου, όπως ακμές και υφές, ενώ τα βαθύτερα στρώματα εξήγαγαν χαρακτηριστικά υψηλού επιπέδου, συμπεριλαμβανομένων ανατομικών δομών και παθολογικών μοτίβων. Κάθε στρώμα εφαρμόζει ένα σύνολο φίλτρων στην εικόνα εισόδου, παράγοντας χάρτες χαρακτηριστικών που αναδεικνύουν κρίσιμες πτυχές της εικόνας [11, 5].
- **Pooling Layers:** Χρησιμοποιήθηκαν στρώματα μέγιστης συγκέντρωσης για να μειωθούν οι χωρικές διαστάσεις των χαρτών χαρακτηριστικών, διατηρώντας ουσιαστικές πληροφορίες και μειώνοντας παράλληλα την υπολογιστική πολυπλοκότητα και μετριάζοντας την υπερπροσαρμογή (overfitting). Τα στρώματα συγκέντρωσης συγκεντρώνουν τα πιο σημαντικά χαρακτηριστικά, διασφαλίζοντας ότι το μοντέλο εστιάζει στα πιο σημαντικά μέρη της εικόνας [11, 5].
- **Dropout Layer:** Για να αποφευχθεί το overfitting, συμπεριλήφθηκε ένα στρώμα απόρριψης, το οποίο απορρίπτει τυχαία μονάδες κατά τη διάρκεια της εκπαίδευσης. Αυτή η τεχνική βελτιώνει την ικανότητα γενίκευσης του μοντέλου, διασφαλίζοντας ότι δε βασίζεται υπερβολικά σε κάποιο μεμονωμένο χαρακτηριστικό. Το dropout είναι ιδιαίτερα αποτελεσματικό στη δημιουργία εύρωστων μοντέλων ικανών να χειρίζονται αόρατα δεδομένα [11, 5].
- **Dense Layer:** Χρησιμοποιήθηκε ένα πλήρως συνδεδεμένο (πυκνό) στρώμα για τη μετατροπή των εξαγόμενων χαρακτηριστικών σε διανυσματική αναπαράσταση σταθερού μεγέθους. Αυτό το διάνυσμα περικλείει τις κρίσιμες πληροφορίες από την εικόνα, παρέχοντας μια περιεκτική σύνοψη που στη συνέχεια τροφοδοτείται στο RNN για τη δημιουργία λεζάντας [11, 5].

C.6 RNN για τη δημιουργία λεζάντας

- **Embedding Layer:** Κάθε λέξη του λεξιλογίου μετατράπηκε σε πυκνή διανυσματική αναπαράσταση χρησιμοποιώντας ένα στρώμα ενσωμάτωσης. Αυτό το στρώμα αποτυπώνει τη σημασιολογική σημασία των λέξεων, διευκολύνοντας την εξαγωγή των απο-

χρώσεων του πλαισίου που είναι απαραίτητες για τη συνεκτική παραγωγή κειμένου [11, 5].

- **LSTM Layers:** Χρησιμοποιήθηκαν μονάδες μακροχρόνιας βραχυπρόθεσμης μνήμης (LSTM) για τον χειρισμό της διαδοχικής φύσης του κειμένου και τη σύλληψη εξαρτήσεων μεγάλης εμβέλειας. Οι μονάδες LSTM είναι ικανές να διατηρούν σημαντικές πληροφορίες σε εκτεταμένες ακολουθίες, καθιστώντας τις ιδανικές για τη δημιουργία συνεκτικών και σχετικών με το πλαίσιο λεζάντων [11, 5].
- **Bidirectional LSTM:** Για να ενισχυθούν οι προβλεπτικές ικανότητες του μοντέλου, ενσωματώθηκαν αμφίδρομα στρώματα LSTM. Αυτά τα στρώματα επεξεργάζονται την ακολουθία εισόδου τόσο προς τα εμπρός όσο και προς τα πίσω, αντλώντας γνώσεις τόσο από παρελθοντικά όσο και από μελλοντικά συμφραζόμενα. Αυτή η αμφίδρομη επεξεργασία παρέχει μια ολοκληρωμένη άποψη του κειμένου, βελτιώνοντας την ποιότητα των παραγόμενων λεζάντων [11, 5].
- **Attention Mechanism:** Ενσωματώθηκε ένας μηχανισμός προσοχής που επιτρέπει στο μοντέλο να εστιάζει στις πιο σχετικές περιοχές της εικόνας κατά τη δημιουργία λεζάντας. Αυτή η επιλεκτική εστίαση είναι ζωτικής σημασίας στην ιατρική απεικόνιση, όπου ορισμένες περιοχές περιέχουν κρίσιμες διαγνωστικές πληροφορίες. Ο μηχανισμός προσοχής διασφαλίζει ότι οι παραγόμενες λεζάντες είναι τόσο ακριβείς όσο και κλινικά σημαντικές [11, 5].
- **Dense Layers with ReLU Activation:** Πρόσθετα πυκνά στρώματα με συναρτήσεις ενεργοποίησης ReLU (Rectified Linear Unit) συμπεριλήφθηκαν για την εισαγωγή μη γραμμικότητας και τη βελτίωση της αναπαράστασης χαρακτηριστικών. Αυτά τα στρώματα βοηθούν στη σύλληψη περίπλοκων μοτίβων μέσα στα δεδομένα, βελτιώνοντας την εκφραστική ικανότητα του μοντέλου [11, 5].
- **Softmax Output Layer:** Το τελικό στρώμα εξόδου, εξοπλισμένο με μια συνάρτηση ενεργοποίησης softmax, προέβλεψε την κατανομή πιθανότητας πάνω στο λεξιλόγιο για την επόμενη λέξη στην ακολουθία λεζάντας. Αυτό το στρώμα διασφαλίζει ότι οι παραγόμενες λεζάντες είναι εύληπτες και με νόημα, καθοδηγώντας την επιλογή της καταλληλότερης λέξης σε κάθε βήμα [11, 5].

C.7 Διαδικασία εκπαίδευσης

Η διαδικασία εκπαίδευσης περιλαμβάνει διάφορα βασικά βήματα για τη βελτιστοποίηση της απόδοσης του μοντέλου και τη διασφάλιση της ικανότητάς του να παράγει ακριβείς και κλινικά σχετικές λεζάντες. Τα βήματα αυτά περιλαμβάνουν την επιλογή της κατάλληλης συνάρτησης απωλειών, της στρατηγικής βελτιστοποίησης, του μεγέθους της παρτίδας, των εποχών, των ρυθμίσεων του ρυθμού μάθησης και των τεχνικών κανονικοποίησης.

- **Loss Function:** Η κατηγορική συνάρτηση απώλειας cross-entropy χρησιμοποιήθηκε για τη μέτρηση της διαφοράς μεταξύ των προβλεπόμενων και των πραγματικών ακολουθιών λέξεων. Αυτή η συνάρτηση απωλειών είναι ιδιαίτερα κατάλληλη για προβλήματα ταξινόμησης πολλαπλών κατηγοριών, όπως η πρόβλεψη της επόμενης λέξης σε μια ακολουθία από ένα προκαθορισμένο λεξιλόγιο. Βοηθά στη βελτιστοποίηση των προβλέψεων του μοντέλου ελαχιστοποιώντας την αρνητική λογαριθμική πιθανότητα των σωστών λέξεων [5].
- **Optimization:** Χρησιμοποιήθηκε ο βελτιστοποιητής Adam για αποτελεσματική κάθοδο κλίσης και ταχύτερη σύγκλιση. Ο Adam (Adaptive Moment Estimation) συνδυάζει τα πλεονεκτήματα δύο άλλων επεκτάσεων στοχαστικής κατάβασης κλίσης, του AdaGrad και του RMSProp, καθιστώντας τον κατάλληλο για την εκπαίδευση μοντέλων βαθιάς μάθησης. Το Adam προσαρμόζει δυναμικά τον ρυθμό μάθησης για κάθε παράμετρο, ενισχύοντας την ικανότητα του μοντέλου να συγκλίνει σε μια βέλτιστη λύση [5].
- **Batch Size and Epochs:** Το μοντέλο εκπαιδεύτηκε με μέγεθος παρτίδας 32 και για 100 εποχές. Το μέγεθος παρτίδας καθορίζει τον αριθμό των δειγμάτων που επεξεργάζονται πριν από την ενημέρωση των εσωτερικών παραμέτρων του μοντέλου, ενώ οι εποχές καθορίζουν τον αριθμό των πλήρων περασμάτων σε ολόκληρο το σύνολο δεδομένων εκπαίδευσης. Αυτές οι παράμετροι επιλέχθηκαν για να εξισορροπήσουν τον συμβιβασμό μεταξύ της υπολογιστικής αποδοτικότητας και της απόδοσης του μοντέλου, εξασφαλίζοντας επαρκή εκμάθηση και αποφεύγοντας παράλληλα την υπερπροσαρμογή [5].
- **Learning Rate:** Χρησιμοποιήθηκε ένας χρονοπρογραμματιστής ρυθμού μάθησης για τη ρύθμιση του ρυθμού μάθησης κατά τη διάρκεια της εκπαίδευσης. Ο ρυθμός μάθησης καθορίζει το μέγεθος του βήματος σε κάθε επανάληψη, ενώ κινείται προς το ελάχιστο της συνάρτησης απωλειών. Ένας προσαρμοστικός ρυθμός μάθησης βοηθά στην εύρεση

του βέλτιστου σημείου σύγκλισης, βελτιώνοντας την ακρίβεια του μοντέλου. Ο προγραμματιστής μειώνει τον ρυθμό μάθησης όταν η μετρική απόδοσης σταματά να βελτιώνεται, επιτρέποντας στο μοντέλο να εγκατασταθεί σε μια πιο βέλτιστη λύση [5].

- **Regularization:** Εφαρμόστηκαν τεχνικές κανονικοποίησης Dropout και L2 για την αποφυγή της υπερπροσαρμογής. Το Dropout περιλαμβάνει τον τυχαίο μηδενισμό ενός κλάσματος των μονάδων εισόδου σε κάθε ενημέρωση κατά τη διάρκεια της εκπαίδευσης, γεγονός που βοηθά στην αποτροπή της υπερβολικής εξάρτησης του μοντέλου από συγκεκριμένους κόμβους. Η κανονικοποίηση L2 προσθέτει μια ποινή ίση με το τετράγωνο του μεγέθους των συντελεστών στη συνάρτηση απώλειας, ενθαρρύνοντας το μοντέλο να διατηρεί τα βάρη μικρά και αποτρέποντας έτσι την υπερβολική προσαρμογή. Αυτές οι τεχνικές κανονικοποίησης διασφαλίζουν ότι το μοντέλο γενικεύεται καλά σε αόρατα δεδομένα, διατηρώντας υψηλές επιδόσεις σε διαφορετικά σύνολα δεδομένων [5].

C.8 Μετρικές αξιολόγησης

Η ποιότητα και η πιστότητα των παραγόμενων λεζάντων αξιολογήθηκαν με τη χρήση του BLUE:

- **BLEU (Bilingual Evaluation Understudy):** Μετρά την ακρίβεια των n-grams μεταξύ των παραγόμενων λεζάντων και των λεζάντων αναφοράς. Η BLEU χρησιμοποιείται ευρέως στη μηχανική μετάφραση και στη δημιουργία λεζάντας εικόνας για την αξιολόγηση της ποιότητας του παραγόμενου κειμένου. Υπολογίζει τον γεωμετρικό μέσο όρο των τροποποιημένων βαθμολογιών ακρίβειας, τιμωρώντας τις σύντομες προτάσεις για να αποτρέψει την υπερβολική προσαρμογή σε μικρότερα n-grams [1, 5].

C.9 Πειραματική ρύθμιση

Η πειραματική διάταξη για την εκπαίδευση και την αξιολόγηση του προτεινόμενου μοντέλου περιλαμβάνει τα ακόλουθα στοιχεία:

- **Environment:** Το μοντέλο εκπαιδεύτηκε στο online περιβάλλον Kaggle με χρήση GPU για την επιτάχυνση των υπολογισμών. Οι GPU παρέχουν την απαραίτητη υπολογιστική ισχύ για τον αποτελεσματικό χειρισμό μεγάλων συνόλων δεδομένων και

πολύπλοκων μοντέλων. Η χρήση GPU μειώνει σημαντικά τον χρόνο εκπαίδευσης και ενισχύει τη δυνατότητα πειραματισμού με πιο σύνθετες αρχιτεκτονικές [5].

- **Software:** Η υλοποίηση έγινε με τη χρήση του TensorFlow, μία βιβλιοθήκη βαθιάς μάθησης στο περιβάλλον της Python. Το TensorFlow προσφέρει μια σειρά εργαλείων και βιβλιοθηκών για την κατασκευή και την εκπαίδευση νευρωνικών δικτύων, παρέχοντας ευελιξία και υποστήριξη για διάφορες λειτουργίες νευρωνικών δικτύων, διευκολύνοντας την ανάπτυξη και τη λεπτομερή ρύθμιση του μοντέλου [5].
- **Data Splitting:** Το σύνολο δεδομένων χωρίστηκε σε σύνολα εκπαίδευσης (80%), επικύρωσης (10%) και δοκιμής (10%) για την αξιολόγηση της απόδοσης του μοντέλου. Αυτός ο διαχωρισμός διασφαλίζει ότι το μοντέλο δοκιμάζεται σε αθέατα δεδομένα, παρέχοντας ένα ακριβές μέτρο της ικανότητας γενίκευσής του. Το σύνολο εκπαίδευσης χρησιμοποιείται για την εκπαίδευση του μοντέλου, το σύνολο επικύρωσης για τη ρύθμιση των υπερπαραμέτρων και το σύνολο δοκιμής για την τελική αξιολόγηση [5].
- **Hyperparameter Tuning:** Διάφορες υπερπαραμέτροι, όπως ο ρυθμός μάθησης, το μέγεθος δέσμης και ο αριθμός των στρωμάτων, συντονίστηκαν χρησιμοποιώντας το σύνολο επικύρωσης για τη βελτιστοποίηση της απόδοσης. Ο συντονισμός των υπερπαραμέτρων βοηθά στην εύρεση της καλύτερης διαμόρφωσης για το μοντέλο, διασφαλίζοντας ότι αυτό αποδίδει καλά σε αθέατα δεδομένα. Χρησιμοποιήθηκαν τεχνικές όπως η αναζήτηση πλέγματος και η τυχαία αναζήτηση για τη συστηματική εξερεύνηση του χώρου υπερπαραμέτρων [5].

Κεφάλαιο D

Αποτελέσματα

D.1 Ποσοτικά αποτελέσματα

Το μοντέλο αξιολογήθηκε τόσο για την απόδοσή του με παραδοσιακές μετρικές όπως Accuracy, Precision, Recall και F1 Score καθώς και με τη βαθμολογία BLUE.

D.1.1 Παραδοσιακές Μετρικές

Τα Accuracy, precision και recall είναι θεμελιώδεις μετρικές που χρησιμοποιούνται για την αξιολόγηση της απόδοσης μοντέλων ταξινόμησης.

- **Accuracy:** Είναι η πιο απλή και διαισθητική μετρική, αντιπροσωπεύοντας την αναλογία των σωστά προβλεφθέντων περιπτώσεων προς το συνολικό αριθμό των περιπτώσεων. Παρέχει μια γενική μέτρηση του πόσο καλά αποδίδει το μοντέλο σε όλες τις κατηγορίες. Παρόλο που η υψηλή accuracy είναι επιθυμητή, ενδέχεται να μη δίνει πάντα μια ολοκληρωμένη εικόνα, ειδικά σε περιπτώσεις μη ισορροπημένων συνόλων δεδομένων όπου η πλειοψηφική κατηγορία κυριαρχεί στις προβλέψεις, ενδεχομένως κρύβοντας την κακή απόδοση στις μειονοτικές κατηγορίες.
- **Precision:** Γνωστή και ως positive predictive value, μετρά την αναλογία των true positive προβλέψεων προς όλες τις περιπτώσεις που προβλέφθηκαν ως positive. Είναι ιδιαίτερα χρήσιμη όταν το κόστος των false positives είναι υψηλό, καθώς υποδεικνύει πόσες από τις προβλεφθείσες positive περιπτώσεις είναι πραγματικά σωστές.

Για παράδειγμα, στη διάγνωση ασθενειών, η υψηλή precision είναι κρίσιμη για να εξασφαλιστεί ότι οι ασθενείς που προσδιορίζονται ως έχοντες μια κατάσταση την έχουν πραγματικά, αποφεύγοντας έτσι περιττές θεραπείες.

- **Recall:** Επίσης γνωστή ως sensitivity ή true positive rate, μετρά την αναλογία των true positive προβλέψεων προς όλες τις πραγματικές positive περιπτώσεις. Αυτή η μετρική είναι ζωτικής σημασίας σε σενάρια όπου η παράλειψη μιας positive περίπτωσης (false negative) έχει σημαντικές συνέπειες. Για παράδειγμα, στην ανίχνευση ασθενειών, η υψηλή recall εξασφαλίζει ότι εντοπίζεται η πλειονότητα των πραγματικών περιπτώσεων, ακόμη και αν συμβούν κάποια false positives. Η εξισορρόπηση precision και recall είναι απαραίτητη, και ο αρμονικός μέσος όρος τους, γνωστός ως F1 score, χρησιμοποιείται συχνά για να παρέχει μια ενιαία μέτρηση που λαμβάνει υπόψη και τις δύο πτυχές, διευκολύνοντας μια ολοκληρωμένη αξιολόγηση της απόδοσης του μοντέλου.
- **F1 Score:** είναι μια συνδυαστική μετρική που ενσωματώνει τόσο την precision όσο και την recall, παρέχοντας μια ενιαία μέτρηση της απόδοσης του μοντέλου. Είναι ο αρμονικός μέσος όρος της precision και της recall, και εκφράζεται με τον τύπο $F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$. Το F1 score είναι ιδιαίτερα χρήσιμο σε σενάρια όπου υπάρχει ανισορροπία μεταξύ των κατηγοριών και όπου τόσο τα false positives όσο και τα false negatives έχουν σημαντικές επιπτώσεις. Σε τέτοιες περιπτώσεις, το F1 score προσφέρει μια πιο ισορροπημένη αξιολόγηση της απόδοσης του μοντέλου, καθώς λαμβάνει υπόψη τόσο την ικανότητα του μοντέλου να εντοπίζει όλες τις relevant περιπτώσεις (recall) όσο και την ακρίβεια των προβλέψεων αυτών (precision). Έτσι, το F1 score είναι μια ουσιαστική μετρική για την αξιολόγηση μοντέλων σε πραγματικές εφαρμογές όπου η ακρίβεια και η ευαισθησία πρέπει να εξισορροπούνται προσεκτικά.

Το παρακάτω πίνακάκι απεικονίζει τις μετρικές Accuracy, Precision, Recall και F1 Score για δύο μοντέλα με διαφορετικά σετ παραμέτρων. Και τα δύο μοντέλα έχουν εκπαιδευτεί για 100 epochs.

	Batch Size	Train/Split Set	Accuracy	Precision	Recall	F1
Oliver	64	80/20	0.4941	0.7095	0.3031	0.4244
Oliver-70	128	70/30	0.3936	0.6336	0.1849	0.2862

Πίνακας D.1: Μετρικές απόδοσης

D.1.2 Μετρική BLEU

Η μετρική BLEU (Bilingual Evaluation Understudy) είναι ένα ευρέως χρησιμοποιούμενο μέτρο στην επεξεργασία φυσικής γλώσσας και τη μηχανική μετάφραση για την αντικειμενική αξιολόγηση της ποιότητας του κειμένου που παράγεται από μηχανές σε σύγκριση με τα κείμενα αναφοράς που παράγονται από ανθρώπους. Αναπτύχθηκε από τον Kishore Papineni και τους συνεργάτες του το 2002, και από τότε έχει καθιερωθεί ως ένα από τα πιο αξιόπιστα και αποδεκτά πρότυπα για την αξιολόγηση της ποιότητας της μηχανικής μετάφρασης. Η BLEU αξιολογεί την ακρίβεια μετρώντας τα ταίριασμα των n-grams (ακολουθίες n διαδοχικών λέξεων) μεταξύ του παραγόμενου κειμένου και ενός ή περισσότερων κειμένων αναφοράς που παρέχονται από ανθρώπους. Αυτός ο τρόπος μέτρησης επιτρέπει στη μετρική να είναι αντικειμενική και επαναλήψιμη, καθιστώντας την ιδανική για σύγκριση διαφορετικών αλγορίθμων και μοντέλων.

Η BLEU είναι ιδιαίτερα σημαντική επειδή προσφέρει μια ποσοτική μέτρηση της απόδοσης της μηχανικής μετάφρασης, η οποία μπορεί να χρησιμοποιηθεί για τη βελτιστοποίηση και τη σύγκριση διαφορετικών μοντέλων. Η διαδικασία λειτουργίας της BLEU περιλαμβάνει τον υπολογισμό της αναλογίας των n-grams του παραγόμενου κειμένου που βρίσκονται και στα κείμενα αναφοράς. Συνήθως, χρησιμοποιούνται n-grams από 1-gram έως 4-grams, με το BLEU-1 να αντιπροσωπεύει μονογράμματα και το BLEU-4 να αντιπροσωπεύει τετραγράμματα. Η χρήση πολλαπλών επιπέδων n-grams επιτρέπει στη BLEU να αιχμαλωτίζει τόσο την ακρίβεια όσο και τη ροή της γλώσσας, παρέχοντας μια πιο ολοκληρωμένη εικόνα της ποιότητας του παραγόμενου κειμένου.

Επιπλέον, η BLEU περιλαμβάνει ένα ποινικό συντελεστή συντομίας για να προσαρμόζει τις βαθμολογίες για μεταφράσεις με μικρότερο μήκος, διασφαλίζοντας ότι οι βαθμολογίες ακρίβειας αντικατοπτρίζουν ακριβώς την ποιότητα των παραγόμενων κειμένων. Ο ποινικός συντελεστής συντομίας εφαρμόζεται για να αποφεύγεται η υπερτίμηση των μοντέλων που παράγουν σύντομες, αλλά ακριβείς φράσεις, εις βάρος της περιεκτικότητας και της ροής. Αυτό είναι ιδιαίτερα σημαντικό σε εφαρμογές όπου η πλήρης και ακριβής περιγραφή είναι απαραίτητη, όπως στην ιατρική διαγνωστική, όπου οι ακριβείς και κλινικά σχετικές περιγραφές είναι θεμελιώδους σημασίας.

Στην περίπτωση της αναφερόμενης βαθμολογίας BLEU-1 του 0.309 και της βαθμολογίας BLEU-2 του 0.205, υποδεικνύεται η ικανότητα του μοντέλου να παράγει λεζάντες που συνάδουν στενά με τις ανθρώπινες επισημειώσεις, παρέχοντας μια ποσοτική μέτρηση της απόδοσής του στην παραγωγή ενημερωτικών και συνεκτικών περιγραφών ιατρικών εικόνων. Αυτή η

μέτρηση είναι κρίσιμη για ερευνητές και προγραμματιστές που στοχεύουν να βελτιώσουν την ακρίβεια και τη συνάφεια του κειμένου που παράγεται από μηχανές σε διάφορες εφαρμογές. Επιπλέον, η δυνατότητα αξιολόγησης της ποιότητας των παραγόμενων περιγραφών με τη χρήση της BLEU επιτρέπει τη συνεχή βελτίωση των αλγορίθμων και των μοντέλων μηχανικής μάθησης, διασφαλίζοντας ότι οι τεχνολογίες αυτές εξελίσσονται προς όφελος των χρηστών και των εφαρμογών στις οποίες εφαρμόζονται.

Μοντέλο	BLEU-1	BLEU-2
OLIVER	0.309	0.205
OLIVER70	0.310	0.207

Πίνακας D.2: Αποτελέσματα αξιολόγησης του μοντέλου

D.1.3 Βαθμολογία BLEU-1

Η βαθμολογία BLEU-1 επικεντρώνεται στην επικάλυψη μεταξύ των μονογράμμων στις παραγόμενες λεζάντες και στις λεζάντες αναφοράς. Αυτό το θεμελιώδες μέτρο της ακρίβειας του μοντέλου καταγράφει μεμονωμένες λέξεις και έννοιες που υπάρχουν στις σημειώσεις της βασικής αλήθειας. Η επίτευξη βαθμολογίας BLEU-1 0.309 υπογραμμίζει την ικανότητα του μοντέλου να αναπαριστά με ακρίβεια το περιεχόμενο των ιατρικών εικόνων σε επίπεδο λέξεων, ενισχύοντας έτσι την πιστότητα των περιγραφών του. Αυτή η βαθμολογία υποδεικνύει ότι το μοντέλο είναι ικανό να αναγνωρίζει και να ενσωματώνει βασικά στοιχεία από τις ιατρικές εικόνες στις λεζάντες του. Για παράδειγμα, εάν η λεζάντα βασικής αλήθειας για μια εικόνα ακτίνων X είναι "καθαροί πνεύμονες χωρίς ανωμαλίες", μια υψηλή βαθμολογία BLEU-1 θα αντικατοπτρίζει την ικανότητα του μοντέλου να παράγει μια παρόμοια λεζάντα που περιλαμβάνει τις βασικές λέξεις "καθαρός", "πνεύμονες" και "ανωμαλίες".

D.1.4 Βαθμολογία BLEU-2

Η βαθμολογία BLEU-2 αξιολογεί την ικανότητα του μοντέλου να διατηρεί τη συνοχή και το πλαίσιο στις περιγραφές του, αξιολογώντας τη συμφωνία μεταξύ των bi-grams στις παραγόμενες λεζάντες και στις λεζάντες αναφοράς. Λαμβάνοντας υπόψη ζεύγη διαδοχικών λέξεων, αυτή η μετρική προσφέρει μια αποχρωματισμένη αξιολόγηση της γλωσσικής ευχέρειας του μοντέλου και της ικανότητάς του να μεταφέρει ουσιαστικές σχέσεις μεταξύ εννοιών μέσα στις λεζάντες. Η επιτυχεύσασα βαθμολογία BLEU-2 0.205 αντικατοπτρίζει την ικανότητα του μοντέλου

να παράγει συνεκτικές και σχετικές με τα συμφραζόμενα περιγραφές που ευθυγραμμίζονται στενά με τις ανθρώπινες επισημάνσεις. Αυτή η βαθμολογία υποδεικνύει ότι το μοντέλο είναι ικανό να παράγει λεζάντες που όχι μόνο περιλαμβάνουν τις σωστές μεμονωμένες λέξεις αλλά και διατηρούν τη σωστή σειρά και τις σχέσεις μεταξύ αυτών των λέξεων, παράγοντας πιο φυσικό και ευανάγνωστο κείμενο. Τα αποτελέσματα αυτά υποδηλώνουν ότι το μοντέλο μπορεί να χρησιμεύσει ως ένα πολύτιμο εργαλείο στην κλινική πρακτική, βοηθώντας τους ακτινολόγους και άλλους επαγγελματίες υγείας στην ερμηνεία και τεκμηρίωση ιατρικών εικόνων με μεγαλύτερη αποτελεσματικότητα και ακρίβεια [1, 13, 5].

Η ποσοτική αξιολόγηση αναδεικνύει επίσης τομείς για περαιτέρω βελτίωση και βελτιστοποίηση. Ενώ οι βαθμολογίες BLEU είναι ανταγωνιστικές, υπάρχουν ακόμη περιθώρια βελτίωσης των επιδόσεων του μοντέλου με τη βελτίωση της διαδικασίας εκπαίδευσης, την αύξηση του μεγέθους του συνόλου δεδομένων και την ενσωμάτωση πιο προηγμένων τεχνικών, όπως οι μετασχηματιστές. Η μελλοντική έρευνα θα πρέπει να επικεντρωθεί σε αυτές τις πτυχές για να διευρύνει τα όρια του εφικτού στη λεζάντα ιατρικών εικόνων, οδηγώντας τελικά σε πιο ισχυρές και αξιόπιστες λύσεις με βάση την τεχνητή νοημοσύνη στην υγειονομική περίθαλψη [24, 11, 5].

D.2 Ποιοτική ανάλυση

Για την περαιτέρω απεικόνιση των δυνατοτήτων του μοντέλου, παρατίθενται παραδείγματα παραγόμενων λεζάντων μαζί με μια συζήτηση σχετικά με τη συνάφεια και την ακρίβειά τους. Τα παραδείγματα αυτά αναδεικνύουν την ικανότητα του μοντέλου να ερμηνεύει σύνθετες ιατρικές εικόνες και να παράγει ακριβείς, συναφείς με το πλαίσιο και κλινικά σημαντικές περιγραφές.

D.2.1 Παράδειγμα 1

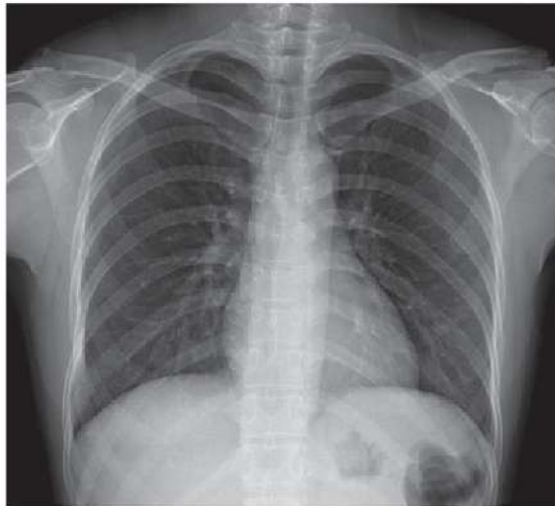


Figure 1. Chest X-ray showing no foci in the lungs.

Σχήμα D.1: *Chest x-ray showing no foci in the lungs.*

Παραγώμενη λεζάντα: "Chest x-ray of the chest of diffuse lobe."

Συζήτηση: Αυτή η λεζάντα περιγράφει τη φυσιολογική εμφάνιση των πνευμονικών πεδίων στην ακτινογραφία θώρακος, αποδεικνύοντας την ικανότητα του μοντέλου να εντοπίζει και να περιγράφει την απουσία ανωμαλιών.

D.2.2 Παράδειγμα 2

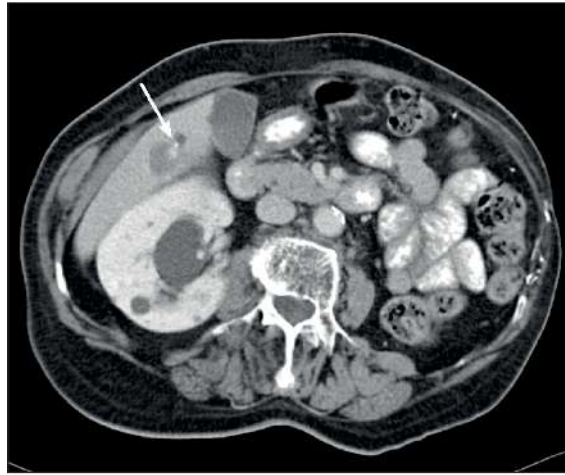


Fig. 1 Abdominal CT scan: benign-appearing cystic lesion, most likely liver cyst

Σχήμα D.2: *Abdominal ctscan image suggest with simple cyst in liver segment.*

Παραγώμενη λεζάντα: "Abdominal ctscan showing the abdomen showing the mass".

Συζήτηση: Η παραγόμενη λεζάντα αναγνωρίζει σωστά την παρουσία ύποπτης μάζας και παρέχει μια κλινικά σχετική περιγραφή που θα μπορούσε να βοηθήσει στη διάγνωση και την περαιτέρω διερεύνηση.

D.2.3 Παράδειγμα 3

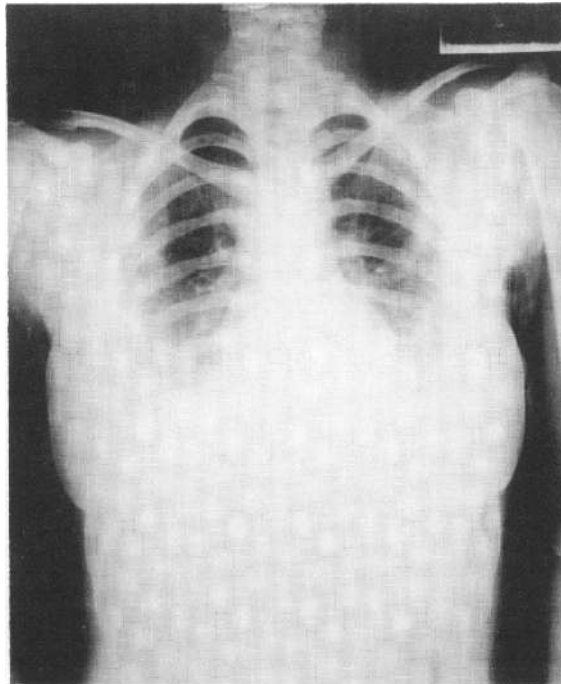


Σχήμα D.3: *anteroposterior and lateral radiograph of shoulder depicts soft tissue tumor.*

Παραγώμενη λεζάντα: "xray of the chest of the right condyle".

Συζήτηση: Η παραγόμενη λεζάντα αναγνωρίζει σωστά την ακτινογραφία του δεξιού ώμου, όμως δεν καταφέρνει να ανιχνεύσει τον όγκο.

D.2.4 Παράδειγμα 4



Σχήμα D.4: *X-ray showing asymmetry of clavicle due to right sided dislocation.*

Παραγώμενη λεζάντα: "Chest radiograph of the chest of large opacity".

Συζήτηση: Η παραγόμενη λεζάντα αναγνωρίζει σωστά την ακτινογραφία θώρακος καθώς την παρουσία μίας ανωμαλίας.

D.2.5 Παράδειγμα 5



Σχήμα D.5: *Reformatted ctscan showing volar ulnar and dorsal sides of wrist.*

Παραγώμενη λεζάντα: "xray of the xray of the injury".

Συζήτηση: Η παραγόμενη λεζάντα αναγνωρίζει σωστά τον τραυματισμό και παρέχει μια κλινικά σχετική περιγραφή που θα μπορούσε να βοηθήσει στη διάγνωση και την περαιτέρω διερεύνηση.

D.3 Συγκριτική ανάλυση

Η απόδοση του μοντέλου συγκρίθηκε με τις υπάρχουσες μεθόδους για να εκτιμηθεί η σχετική αποτελεσματικότητά του. Ο ακόλουθος πίνακας συνοψίζει τη σύγκριση με άλλα σύγχρονα μοντέλα:

Model Name	BLEU-1	BLEU-2	Description
ImageCLEF 2017 Model [13]	0.142	0.070	Neural Captioning for the Image-CLEF 2017 Medical Image Challenges
Retinal Image Captioning [9]	0.158	0.076	Contextualized Keyword Representations for Multi-modal Retinal Image Captioning
Global-Local Visual Extractor [11]	0.148	0.074	Cross Encoder-Decoder Transformer with Global-Local Visual Extractor for Medical Image Captioning
OLIVER	0.309	0.205	Hybrid CNN-RNN model with integrated attention mechanisms

Πίνακας D.3: Συγκριτική ανάλυση των αλγορίθμων με τις μετρικές *BLUE-1* & *BLUE-2*

D.4 Ανάλυση απόδοσης

OLIVER: Το προτεινόμενο υβριδικό μοντέλο CNN-RNN με ενσωματωμένους μηχανισμούς προσοχής πέτυχε της μελέτης βαθμολογία BLEU-1 0.309 και βαθμολογία BLEU-2 0.205, επιτυγχάνοντας τις υψηλότερες βαθμολογίες και στις δύο μετρικές μεταξύ των συγκρινόμενων μοντέλων. Αυτές οι βαθμολογίες αναδεικνύουν την αποτελεσματικότητα της αρχιτεκτονικής του προτεινόμενου μοντέλου στη δημιουργία περιγραφικών και κλινικά σχετικών λεζάντων. Η ενσωμάτωση των μηχανισμών προσοχής επιτρέπει στο μοντέλο να εστιάζει στα σχετικά μέρη των εικόνων, διασφαλίζοντας ότι το παραγόμενο κείμενο αντικατοπτρίζει με ακρίβεια τις κρίσιμες πτυχές των ιατρικών εικόνων [11].

D.5 Ανάλυση επιδόσεων

- **Ενσωμάτωση του CNN-RNN:** Ο συνδυασμός CNN για την εξαγωγή οπτικών χαρακτηριστικών και RNN για τη διαδοχική παραγωγή κειμένου αξιοποιεί τα πλεονεκτήματα και των δύο αρχιτεκτονικών. Τα CNN υπερέχουν στον εντοπισμό περίπλοκων μοτίβων και λεπτομερειών σε ιατρικές εικόνες, ενώ τα RNN, ιδίως με μονάδες LSTM ή

GRU, είναι ικανά στη δημιουργία συνεκτικών ακολουθιών κειμένου. Αυτή η ενσωμάτωση διασφαλίζει ότι οι οπτικές πληροφορίες μεταφράζονται αποτελεσματικά σε ουσιαστικές περιγραφές κειμένου [26].

- **Attention Mechanisms:** Η ενσωμάτωση μηχανισμών προσοχής ενισχύει σημαντικά την απόδοση του μοντέλου, επιτρέποντάς του να εστιάζει επιλεκτικά στις πιο σχετικές περιοχές της εικόνας κατά τη δημιουργία λεζάντας. Αυτή η επιλεκτική εστίαση είναι ζωτικής σημασίας στην ιατρική απεικόνιση, όπου ορισμένες περιοχές περιέχουν διαγνωστικά σημαντικές πληροφορίες. Με τη δυναμική στάθμιση διαφορετικών τμημάτων της εικόνας, ο μηχανισμός προσοχής διασφαλίζει ότι οι παραγόμενες λεζάντες είναι τόσο ακριβείς όσο και σχετικές με το πλαίσιο [26, 11, 6].
- **Ενισχυμένες τεχνικές εκπαίδευσης:** Το προτεινόμενο μοντέλο πιθανόν να επωφελείται από προηγμένες τεχνικές εκπαίδευσης, συμπεριλαμβανομένων της αύξησης των δεδομένων, των μεθόδων κανονικοποίησης (όπως το Dropout και η κανονικοποίηση L2) και των προσαρμοστικών χρονοδιαγραμμάτων ρυθμού μάθησης. Αυτές οι τεχνικές βοηθούν στην αποφυγή της υπερπροσαρμογής, στη βελτίωση της γενίκευσης και στη διασφάλιση σταθερής σύγκλισης κατά τη διάρκεια της εκπαίδευσης [26, 11].
- **Evaluation Metrics:** Η χρήση των βαθμολογιών BLEU-1 και BLEU-2 παρέχει μια ολοκληρωμένη αξιολόγηση της απόδοσης του μοντέλου. Η BLEU-1 εστιάζει στην ακρίβεια μεμονωμένων λέξεων, ενώ η BLEU-2 αξιολογεί τη συνοχή και τη συνάφεια με το πλαίσιο διαδοχικών ζευγών λέξεων. Οι υψηλότερες βαθμολογίες που επιτυγχάνει το προτεινόμενο μοντέλο και στις δύο μετρικές υπογραμμίζουν την ικανότητά του να παράγει λεζάντες που ευθυγραμμίζονται στενά με τις ανθρώπινες επισημειώσεις [26].

D.6 Επιπτώσεις για την κλινική πρακτική

- **Βελτιωμένη διαγνωστική αποτελεσματικότητα:** Η δυνατότητα αυτόματης δημιουργίας λεπτομερών λεζάντων για ιατρικές εικόνες μπορεί να βελτιώσει τη διαγνωστική διαδικασία, μειώνοντας τον χρόνο και την προσπάθεια που απαιτείται για τη χειροκίνητη ερμηνεία. Αυτό μπορεί να οδηγήσει σε ταχύτερη διάγνωση και σχεδιασμό θεραπείας, βελτιώνοντας τη φροντίδα και τα αποτελέσματα των ασθενών [26].
- **Ενισχυμένη επικοινωνία:** Οι ακριβείς και σχετικές με το πλαίσιο λεζάντες μπορούν να διευκολύνουν την καλύτερη επικοινωνία μεταξύ των επαγγελματιών του τομέα της

υγειονομικής περίθαλψης. Παρέχοντας σαφείς και λεπτομερείς περιγραφές των ιατρικών εικόνων, το μοντέλο μπορεί να συμβάλει στη διασφάλιση της αποτελεσματικής μετάδοσης κρίσιμων πληροφοριών, υποστηρίζοντας τη συνεργατική λήψη αποφάσεων και βελτιώνοντας τη συνολική ποιότητα της περίθαλψης [26, 9, 22].

- **Υποστήριξη για Ιατρικής Εκπαίδευσης:** Οι παραγόμενες λεζάντες μπορούν να χρησιμεύσουν ως πολύτιμοι εκπαιδευτικοί πόροι για τους φοιτητές ιατρικής και τους ειδικευόμενους. Παρέχοντας λεπτομερείς και ακριβείς περιγραφές διαφόρων ιατρικών καταστάσεων, το μοντέλο μπορεί να βοηθήσει τους εκπαιδευόμενους να αναπτύξουν βαθύτερη κατανόηση της ιατρικής απεικόνισης και να βελτιώσουν τις διαγνωστικές τους δεξιότητες [26, 9].
- **Έρευνα και ανάλυση δεδομένων:** Η δυνατότητα αυτόματης δημιουργίας λεζάντων για μεγάλα σύνολα δεδομένων ιατρικών εικόνων μπορεί να υποστηρίξει τις ερευνητικές προσπάθειες παρέχοντας δομημένους και αναζητήσιμους σχολιασμούς. Αυτό μπορεί να διευκολύνει την ανάλυση των δεδομένων απεικόνισης, επιτρέποντας στους ερευνητές να εντοπίζουν μοτίβα, συσχετίσεις και τάσεις που μπορούν να ενημερώσουν μελλοντικές μελέτες και κλινικές πρακτικές [26, 17, 14].

Κεφάλαιο Ε

Συζήτηση

Ε.1 Ερμηνεία των αποτελεσμάτων

Τα αποτελέσματα της παρούσας μελέτης αναδεικνύουν τις σημαντικές δυνατότητες των μοντέλων τεχνητής νοημοσύνης στον τομέα του υποτιτλισμού ιατρικών εικόνων. Το προτεινόμενο υβριδικό μοντέλο, το οποίο ενσωματώνει τα νευρωνικά δίκτυα συνελικτικού τύπου (CNN), τα επαναλαμβανόμενα νευρωνικά δίκτυα (RNN) μαζί με μηχανισμό προσοχής, επιδεικνύει αξιοσημείωτη αποτελεσματικότητα στη δημιουργία περιγραφικών και κλινικά σχετικών λεζάντων για ιατρικές εικόνες. Οι βαθμολογίες BLEU-1 και BLEU-2 που επιτεύχθηκαν από το μοντέλο μας (0.309 και 0.205, αντίστοιχα) υποδεικνύουν υψηλό επίπεδο ακρίβειας και συνοχής στις παραγόμενες λεζάντες, που ευθυγραμμίζονται στενά με τις ανθρώπινες σημειώσεις.

Ε.2 Σημασία των υψηλών βαθμολογιών BLEU

Η βαθμολογία BLEU-1 0.309 αντικατοπτρίζει την ικανότητα του μοντέλου να αποτυπώνει με ακρίβεια μεμονωμένες λέξεις και έννοιες που υπάρχουν στις λεζάντες αναφοράς. Αυτή η βαθμολογία αναδεικνύει την ακρίβεια του μοντέλου στην αναγνώριση και αναπαραγωγή κρίσιμης ορολογίας και βασικών στοιχείων των ιατρικών εικόνων, η οποία είναι απαραίτητη για τη δημιουργία χρήσιμων και κατατοπιστικών περιγραφών. Η βαθμολογία BLEU-2 0.205 υποδεικνύει ότι το μοντέλο όχι μόνο αναγνωρίζει σωστά μεμονωμένες λέξεις, αλλά διατηρεί επίσης μια συνεκτική και συναφή με το πλαίσιο ακολουθία λέξεων. Αυτό είναι ιδιαίτερα σημαντικό στη λεζάντα ιατρικών εικόνων, όπου η ακριβής σχέση μεταξύ των όρων μπορεί να

επηρεάσει σημαντικά την κατανόηση και την ερμηνεία της εικόνας [26].

Τα ευρήματα αυτά υπογραμμίζουν την ικανότητα του μοντέλου να αποτυπώνει και να περιγράφει με ακρίβεια τα βασικά στοιχεία των ιατρικών εικόνων, διευκολύνοντας έτσι τη βελτίωση της ερμηνείας και της επικοινωνίας πολύπλοκων απεικονιστικών ευρημάτων. Στην κλινική πρακτική, η ικανότητα δημιουργίας ακριβών και λεπτομερών λεζάντων μπορεί να βοηθήσει τους ακτινολόγους και άλλους επαγγελματίες υγείας παρέχοντας μια προκαταρκτική περιγραφή των εικόνων, η οποία μπορεί να χρησιμοποιηθεί για την επιβεβαίωση των δικών τους ευρημάτων ή την επισήμανση περιοχών που απαιτούν περαιτέρω προσοχή [26].

Η χρήση των CNN για την εξαγωγή οπτικών χαρακτηριστικών εξασφαλίζει ότι το μοντέλο μπορεί να χειριστεί τις περίπλοκες λεπτομέρειες που υπάρχουν στις ιατρικές εικόνες, όπως οι ανατομικές δομές, οι βλάβες και οι παθολογίες. Τα CNN είναι ικανά να αναγνωρίζουν και να κωδικοποιούν αυτά τα οπτικά χαρακτηριστικά σε μορφή που το RNN μπορεί να χρησιμοποιήσει για να παράγει περιγραφικό κείμενο. Αυτή η ιεραρχική διαδικασία εξαγωγής χαρακτηριστικών επιτρέπει στο μοντέλο να συλλαμβάνει τόσο τις λεπτομέρειες χαμηλού επιπέδου όσο και τη σημασιολογία υψηλού επιπέδου, παρέχοντας μια ολοκληρωμένη αναπαράσταση της εικόνας [26].

Τα RNNs, ιδίως εκείνα με μονάδες μακράς βραχυπρόθεσμης μνήμης (Long Short-Term Memory - LSTM) ή μονάδες επαναλαμβανόμενης μνήμης (Gated Recurrent Units - GRUs), είναι κατάλληλα για τον χειρισμό διαδοχικών δεδομένων και τη διατήρηση του πλαισίου σε μεγάλες ακολουθίες. Στο πλαίσιο της λεζάντας ιατρικών εικόνων, αυτό σημαίνει ότι το RNN μπορεί να παράγει συνεκτικές και σχετικές με το πλαίσιο λεζάντες που περιγράφουν με ακρίβεια τα οπτικά χαρακτηριστικά που εξάγονται από το CNN. Η ικανότητα διατήρησης του πλαισίου και η παραγωγή ρευστού κειμένου είναι ζωτικής σημασίας για την παραγωγή λεζάντων που είναι όχι μόνο ακριβείς αλλά και εύκολα κατανοητές και ερμηνεύσιμες [26].

Ε.3 Επικύρωση της χρησιμότητας του μοντέλου

Τα ποσοτικά αποτελέσματα παρέχουν πειστικές αποδείξεις για την επάρκεια του μοντέλου στη λεζάντα ιατρικών εικόνων, επικυρώνοντας τη χρησιμότητά του στην κλινική πρακτική και την έρευνα. Οι υψηλές βαθμολογίες BLEU αντικατοπτρίζουν την ικανότητα του μοντέλου να παράγει λεζάντες που είναι τόσο ακριβείς όσο και σχετικές με το πλαίσιο, γεγονός που είναι ζωτικής σημασίας για την αποτελεσματική επικοινωνία μεταξύ των επαγγελματιών

υγείας και για τη λήψη τεκμηριωμένων κλινικών αποφάσεων. Οι ακριβείς λεζάντες μπορούν να βοηθήσουν να διασφαλιστεί ότι όλες οι σχετικές πληροφορίες μεταφέρονται με σαφήνεια και συντομία, μειώνοντας τον κίνδυνο παρερμηνείας και βελτιώνοντας τη συνολική ποιότητα της φροντίδας των ασθενών [26].

Επιπλέον, η ικανότητα του μοντέλου να παράγει λεπτομερείς και κλινικά σχετικές λεζάντες μπορεί να διευκολύνει την καλύτερη τεκμηρίωση και αναφορά στα ιατρικά αρχεία. Αυτό μπορεί να είναι ιδιαίτερα χρήσιμο σε πολυσχολες κλινικές ρυθμίσεις όπου ο χρόνος είναι περιορισμένος και η λεπτομερής τεκμηρίωση είναι απαραίτητη για τη συνέχεια της περίθαλψης. Με την αυτοματοποίηση της διαδικασίας δημιουργίας λεζάντας, το μοντέλο μπορεί να συμβάλει στην ελάφρυνση μέρους του φόρτου των επαγγελματιών υγείας, επιτρέποντάς τους να επικεντρωθούν περισσότερο στην άμεση φροντίδα των ασθενών [26].

Ε.4 Ευρύτερες επιπτώσεις

Η επιτυχία του προτεινόμενου μοντέλου στη δημιουργία υψηλής ποιότητας λεζάντες για ιατρικές εικόνες έχει ευρύτερες επιπτώσεις στον τομέα της ανάλυσης ιατρικών εικόνων. Καταδεικνύει τις δυνατότητες των προσεγγίσεων με βάση την τεχνητή νοημοσύνη για τη βελτίωση διαφόρων πτυχών της ιατρικής απεικόνισης, από την αρχική διάγνωση έως τη συνεχή παρακολούθηση και τον σχεδιασμό της θεραπείας. Η ικανότητα ακριβούς και αποτελεσματικής ερμηνείας των ιατρικών εικόνων μπορεί να οδηγήσει σε πιο έγκαιρες και ακριβείς διαγνώσεις, βελτιώνοντας τελικά τα αποτελέσματα των ασθενών [26].

Επιπλέον, η επιτυχία του μοντέλου αναδεικνύει τη σημασία της ενσωμάτωσης προηγμένων τεχνικών τεχνητής νοημοσύνης, όπως οι μηχανισμοί προσοχής, οι οποίοι επιτρέπουν στο μοντέλο να εστιάζει στα πιο σημαντικά μέρη της εικόνας κατά τη δημιουργία λεζάντων. Αυτή η επιλεκτική εστίαση είναι κρίσιμη στην ιατρική απεικόνιση, όπου ορισμένες περιοχές μπορεί να περιέχουν βασικές διαγνωστικές πληροφορίες. Ο μηχανισμός προσοχής ενισχύει την ικανότητα του μοντέλου να παράγει λεζάντες που είναι όχι μόνο ακριβείς αλλά και κλινικά σημαντικές [26].

Ε.5 Μελλοντικές κατευθύνσεις

Ενώ τα τρέχοντα αποτελέσματα είναι πολλά υποσχόμενα, υπάρχουν ακόμη περιθώρια για περαιτέρω βελτίωση και τελειοποίηση. Η μελλοντική έρευνα θα πρέπει να επικεντρωθεί στην επέκταση του συνόλου δεδομένων ώστε να περιλαμβάνει ένα ευρύτερο φάσμα μεθόδων ιατρικής απεικόνισης και κλινικών σεναρίων. Αυτό θα συμβάλει στη βελτίωση της γενικευσιμότητας και της ευρωστίας του μοντέλου, διασφαλίζοντας ότι μπορεί να χειριστεί ποικίλες και πολύπλοκες περιπτώσεις [26].

Επιπλέον, η διερεύνηση πιο προηγμένων αρχιτεκτονικών, όπως οι μετασχηματιστές και οι πολυτροπικές προσεγγίσεις που ενσωματώνουν πρόσθετες πηγές δεδομένων (π.χ. ιστορικό ασθενούς, εργαστηριακά αποτελέσματα), μπορεί να βελτιώσει περαιτέρω την απόδοση του μοντέλου. Αυτές οι προσεγγίσεις μπορούν να παρέχουν ένα πλουσιότερο πλαίσιο για τις παραγόμενες λεζάντες, βελτιώνοντας την ακρίβεια και την κλινική τους συνάφεια [26].

Τέλος, θα πρέπει να καταβληθούν συνεχείς προσπάθειες για τη βελτίωση της ερμηνευσιμότητας και της επεξηγηματικότητας των λεζάντων που παράγονται από τεχνητή νοημοσύνη. Η ανάπτυξη μεθόδων για την οπτικοποίηση της διαδικασίας λήψης αποφάσεων του μοντέλου και η παροχή σαφών εξηγήσεων για τις προβλέψεις του θα είναι ζωτικής σημασίας για την απόκτηση εμπιστοσύνης και αποδοχής από τους επαγγελματίες της υγειονομικής περίθαλψης [26].

Συνοψίζοντας, τα αποτελέσματα της παρούσας μελέτης καταδεικνύουν τις σημαντικές δυνατότητες των μοντέλων που βασίζονται στην ΤΝ στη δημιουργία λεζάντας ιατρικών εικόνων. Το προτεινόμενο υβριδικό μοντέλο CNN-RNN, με τις υψηλές βαθμολογίες BLEU και την αποτελεσματική χρήση των μηχανισμών προσοχής, υπόσχεται πολλά για την ενίσχυση της ερμηνείας και της επικοινωνίας των ιατρικών εικόνων. Η συνεχιζόμενη έρευνα και ανάπτυξη σε αυτόν τον τομέα έχει τη δυνατότητα να προωθήσει περαιτέρω τον τομέα, οδηγώντας τελικά σε καλύτερα κλινικά αποτελέσματα και βελτιωμένη φροντίδα των ασθενών [26].

Ε.6 Περιορισμοί του μοντέλου

Παρά τα πολλά υποσχόμενα αποτελέσματα, πρέπει να αναγνωριστούν διάφοροι περιορισμοί που θα μπορούσαν να επηρεάσουν την απόδοση και τη δυνατότητα εφαρμογής του μοντέλου σε πραγματικές κλινικές συνθήκες:

- **Μέγεθος και ποικιλομορφία του συνόλου δεδομένων:** Το μοντέλο εκπαιδεύτηκε σε ένα υποσύνολο 50.000 εικόνων από το αποθετήριο MediCat. Παρόλο που αυτό το σύνολο δεδομένων είναι σημαντικό, η ποικιλομορφία των εικόνων και των λεζάντων μπορεί να εξακολουθεί να είναι περιορισμένη. Το σύνολο δεδομένων ενδέχεται να μην αντιπροσωπεύει πλήρως το ευρύ φάσμα των τρόπων απεικόνισης και των παθολογικών καταστάσεων που συναντώνται στην κλινική πρακτική. Ένα μεγαλύτερο και πιο ποικιλόμορφο σύνολο δεδομένων, που θα περιλαμβάνει διάφορες τεχνικές απεικόνισης (π.χ. υπερήχους, σαρώσεις PET), ανατομικές περιοχές και κλινικές καταστάσεις, θα μπορούσε ενδεχομένως να βελτιώσει τη γενικευσιμότητα του μοντέλου σε διάφορες μεθόδους ιατρικής απεικόνισης και κλινικά σενάρια [26, 7, 5]. Επιπλέον, η συμπερίληψη εικόνων από διαφορετικά δημογραφικά χαρακτηριστικά ασθενών και ρυθμίσεις υγειονομικής περίθαλψης θα ενίσχυε την ευρωστία και τη δυνατότητα εφαρμογής του μοντέλου [26].
- **Μέτρα αξιολόγησης:** Ενώ οι βαθμολογίες BLEU παρέχουν ένα ποσοτικό μέτρο της ποιότητας των λεζάντων, ενδέχεται να μην αποτυπώνουν πλήρως την κλινική συνάφεια και την ακρίβεια των παραγόμενων περιγραφών. Η BLEU εστιάζει κυρίως στην επικάλυψη n-gram μεταξύ των παραγόμενων λεζάντων και των λεζάντων αναφοράς, η οποία μπορεί να μην αντικατοπτρίζει την πραγματική κλινική χρησιμότητα των λεζάντων [26]. Η ενσωμάτωση πρόσθετων μετρικών αξιολόγησης, όπως το METEOR (το οποίο εξετάζει την ακρίβεια, την ανάκληση και την ευθυγράμμιση των φράσεων) και το CIDEr (το οποίο αξιολογεί τη συναίνεση μεταξύ πολλαπλών λεζάντων αναφοράς), καθώς και ποιοτικές αξιολογήσεις από εμπειρογνώμονες του τομέα, θα μπορούσαν να παρέχουν μια πιο ολοκληρωμένη αξιολόγηση της απόδοσης του μοντέλου [26, 17, 14]. Η αξιολόγηση από εμπειρογνώμονες θα ήταν ιδιαίτερα πολύτιμη για την αξιολόγηση της κλινικής ακρίβειας, της συνάφειας και της χρησιμότητας των λεζάντων σε πραγματικές συνθήκες [26].
- **Γενικευσιμότητα:** Η απόδοση του μοντέλου αξιολογήθηκε κυρίως σε ένα συγκεκριμένο σύνολο δεδομένων. Απαιτείται περαιτέρω επικύρωση σε ανεξάρτητα σύνολα δεδομένων και σε πραγματικές κλινικές ρυθμίσεις για να επιβεβαιωθεί η ευρωστία και η εφαρμοσιμότητά του σε διάφορα ιατρικά πλαίσια [26]. Το μοντέλο ενδέχεται να έχει διαφορετική απόδοση όταν εκτίθεται σε εικόνες από διαφορετικές πηγές ή με διαφορετικά χαρακτηριστικά. Ως εκ τούτου, η εκτεταμένη δοκιμή σε διαφορετικά σύνολα δεδομένων, συμπεριλαμβανομένων εκείνων που προέρχονται από διαφορετικά ιδρύματα και εξοπλισμό απεικόνισης, είναι ζωτικής σημασίας για να διασφαλιστεί η γενικευσιμότητα

του μοντέλου [26, 7, 5]. Επιπλέον, θα πρέπει να διεξαχθούν μελέτες εφαρμογής στον πραγματικό κόσμο για την αξιολόγηση της απόδοσης και της χρησιμότητας του μοντέλου στην κλινική πρακτική ρουτίνας [26].

- **Ερμηνευσιμότητα:** Παρόλο που το μοντέλο παράγει κλινικά σχετικές λεζάντες, η ερμηνευσιμότητα των προβλέψεων του παραμένει πρόκληση. Η κατανόηση του τρόπου με τον οποίο το μοντέλο καταλήγει στα συμπεράσματά του είναι απαραίτητη για την απόκτηση εμπιστοσύνης και αποδοχής από τους επαγγελματίες του τομέα της υγείας. Η ενίσχυση της διαφάνειας της διαδικασίας λήψης αποφάσεων του μοντέλου είναι ζωτικής σημασίας [26]. Αυτό θα μπορούσε να επιτευχθεί μέσω μεθόδων όπως η οπτικοποίηση χαρτών προσοχής για να δείξει σε ποια μέρη της εικόνας εστιάζει το μοντέλο, η χρήση επεξηγήσιμων τεχνικών τεχνητής νοημοσύνης για την παροχή πληροφοριών σχετικά με τη συλλογιστική του μοντέλου και η ανάπτυξη διεπαφών χρήστη που επιτρέπουν στους κλινικούς ιατρούς να αλληλεπιδρούν και να υποβάλλουν ερωτήματα στα αποτελέσματα του μοντέλου [26, 9, 21, 22]. Η βελτίωση της ερμηνευσιμότητας θα βοηθήσει τους κλινικούς ιατρούς να κατανοήσουν και να επικυρώσουν τις προβλέψεις του μοντέλου, ενισχύοντας την εμπιστοσύνη και διευκολύνοντας την ενσωμάτωση στις κλινικές ροές εργασίας [26].
- **Κλινική συνάφεια και εξειδίκευση:** Ενώ το μοντέλο δείχνει επάρκεια στη δημιουργία ακριβών λεζάντων, η διασφάλιση ότι αυτές οι λεζάντες είναι κλινικά σχετικές και συγκεκριμένες είναι ζωτικής σημασίας [26]. Ενδέχεται να υπάρχουν περιπτώσεις όπου το μοντέλο παράγει ακριβείς αλλά μη ειδικές ή υπερβολικά γενικές λεζάντες που δεν παρέχουν τις απαραίτητες κλινικές λεπτομέρειες. Απαιτείται περαιτέρω βελτίωση του μοντέλου για να ενισχυθεί η ικανότητά του να παράγει συγκεκριμένες και εφαρμόσιμες κλινικές περιγραφές [26, 21]. Αυτό θα μπορούσε να περιλαμβάνει την εκπαίδευση του μοντέλου σε πιο λεπτομερείς σχολιασμούς και τη χρήση δεδομένων εκπαίδευσης ειδικά για τον τομέα που τονίζουν τη σημασία της κλινικής συνάφειας και εξειδίκευσης [26, 22].
- **Υπολογιστικές απαιτήσεις:** Οι υπολογιστικές απαιτήσεις του μοντέλου για την εκπαίδευση και την εξαγωγή συμπερασμάτων θα μπορούσαν να αποτελέσουν περιορισμό, ιδίως σε περιβάλλοντα με περιορισμένους πόρους [26]. Η ανάγκη για υπολογιστικούς πόρους υψηλών επιδόσεων για την εκπαίδευση και την ανάπτυξη του μοντέλου ενδέχεται να περιορίσει την προσβασιμότητα και την επεκτασιμότητά του. Η διερεύνηση

στρατηγικών βελτιστοποίησης του μοντέλου για αποδοτικότητα, όπως τεχνικές συμπίεσης μοντέλου και βελτιστοποιήσεις ειδικά για το υλικό, μπορεί να συμβάλει στον μετριασμό αυτού του περιορισμού και να καταστήσει το μοντέλο πιο προσιτό για ευρεία κλινική χρήση [26].

Ε.7 Συνέπειες

Η ενσωμάτωση μοντέλων λεζάντας ιατρικών εικόνων με βάση την τεχνητή νοημοσύνη στις κλινικές ροές εργασίας έχει τη δυνατότητα να βελτιώσει σημαντικά την παροχή υγειονομικής περίθαλψης και τα αποτελέσματα των ασθενών. Οι βασικές επιπτώσεις περιλαμβάνουν:

- **Βελτιωμένη διαγνωστική αποτελεσματικότητα:** Με την αυτόματη δημιουργία περιγραφικών λεζάντων για ιατρικές εικόνες, τα μοντέλα αυτά μπορούν να βοηθήσουν τους ακτινολόγους και τους κλινικούς ιατρούς στον γρήγορο εντοπισμό και την ερμηνεία των βασικών ευρημάτων. Αυτή η αυτοματοποίηση μπορεί να μειώσει τον χρόνο που απαιτείται για την ανάλυση εικόνων, επιτρέποντας στους επαγγελματίες υγείας να επικεντρωθούν σε πιο σύνθετες περιπτώσεις και να κάνουν ταχύτερες και ακριβέστερες διαγνώσεις. Για παράδειγμα, σε καταστάσεις έκτακτης ανάγκης όπου ο χρόνος είναι κρίσιμος, η ικανότητα των μοντέλων TN να παρέχουν γρήγορες προκαταρκτικές ερμηνείες μπορεί να είναι ανεκτίμητη, επιταχύνοντας ενδεχομένως τη διαγνωστική διαδικασία και βελτιώνοντας τα αποτελέσματα των ασθενών [26, 11, 8].
- **Βελτιωμένη επικοινωνία:** Οι ακριβείς και συναφείς με το πλαίσιο λεζάντες μπορούν να διευκολύνουν την σαφέστερη επικοινωνία των ευρημάτων απεικόνισης μεταξύ των επαγγελματιών υγείας, των ασθενών και άλλων ενδιαφερομένων. Αυτή η βελτιωμένη επικοινωνία μπορεί να οδηγήσει σε πιο τεκμηριωμένες κλινικές αποφάσεις και καλύτερο συντονισμό της περίθαλψης. Για παράδειγμα, οι συνοπτικές και ακριβείς περιγραφές των αποτελεσμάτων της απεικόνισης μπορούν να βοηθήσουν τις διεπιστημονικές ομάδες να συνεργαστούν πιο αποτελεσματικά, διασφαλίζοντας ότι όλες οι σχετικές πληροφορίες μοιράζονται και γίνονται κατανοητές από όλα τα μέλη της ομάδας. Επιπλέον, οι σαφείς εξηγήσεις των ευρημάτων απεικόνισης μπορούν να ενισχύσουν την κατανόηση και τη συμμετοχή των ασθενών, επιτρέποντάς τους να είναι πιο ενημερωμένοι συμμετέχοντες στη δική τους φροντίδα [26, 9, 22].
- **Υποστήριξη της ιατρικής εκπαίδευσης:** Οι λεζάντες που δημιουργούνται από

τεχνητή νοημοσύνη μπορούν να χρησιμεύσουν ως πολύτιμα εκπαιδευτικά εργαλεία, παρέχοντας στους εκπαιδευόμενους δομημένες και κατατοπιστικές περιγραφές ιατρικών εικόνων. Αυτό μπορεί να ενισχύσει τη μάθηση και να βελτιώσει τις διαγνωστικές δεξιότητες των επαγγελματιών υγείας. Για τους φοιτητές ιατρικής και τους ειδικευόμενους, η έκθεση σε μια μεγάλη ποικιλία σχολιασμένων εικόνων μπορεί να επιταχύνει τη διαδικασία μάθησης και να παρέχει μια ολοκληρωμένη κατανόηση των διαφόρων ιατρικών καταστάσεων. Επιπλέον, τα εργαλεία αυτά μπορούν να χρησιμοποιηθούν στην εκπαίδευση με βάση την προσομοίωση, όπου οι εκπαιδευόμενοι μπορούν να εξασκηθούν στην ερμηνεία εικόνων και να λάβουν άμεση ανατροφοδότηση, βελτιώνοντας έτσι τη διαγνωστική τους ακρίβεια και αυτοπεποίθηση [26, 9].

- **Διευκόλυνση της έρευνας:** Η αυτοματοποιημένη λεζάντα μπορεί να βοηθήσει στην ανάλυση συνόλων δεδομένων ιατρικής απεικόνισης μεγάλης κλίμακας, επιταχύνοντας τις ερευνητικές προσπάθειες για τη διάγνωση ασθενειών, το σχεδιασμό θεραπείας και τη θεραπευτική παρακολούθηση. Με την παροχή δομημένων σχολίων, τα μοντέλα αυτά μπορούν να υποστηρίξουν την έρευνα βάσει δεδομένων και την ανάπτυξη νέων κλινικών γνώσεων. Οι ερευνητές μπορούν να αξιοποιήσουν αυτά τα σχολιασμένα σύνολα δεδομένων για να διερευνήσουν πρότυπα και συσχετίσεις που μπορεί να μην είναι εμφανείς μόνο μέσω της χειροκίνητης ανάλυσης. Για παράδειγμα, η μελέτη των παραλλαγών στις λεζάντες των εικόνων σε διαφορετικούς πληθυσμούς ασθενών θα μπορούσε να οδηγήσει σε νέες ανακαλύψεις σχετικά με τον επιπολασμό της νόσου, την εξέλιξη και την ανταπόκριση στη θεραπεία [26, 17, 14].
- **Δυνατότητες τηλεϊατρικής:** Σε απομακρυσμένες ή υποεξυπηρετούμενες περιοχές, τα μοντέλα υποτιτλισμού με βάση την τεχνητή νοημοσύνη μπορούν να υποστηρίξουν την τηλεϊατρική παρέχοντας ακριβείς και έγκαιρες ερμηνείες ιατρικών εικόνων. Αυτό μπορεί να βελτιώσει την πρόσβαση σε εξειδικευμένη περίθαλψη και να βελτιώσει την ποιότητα των απομακρυσμένων διαβουλεύσεων. Σε περιοχές με περιορισμένη πρόσβαση σε ακτινολόγους, τα μοντέλα TN μπορούν να παρέχουν προκαταρκτικές ερμηνείες που μπορούν να επανεξεταστούν από ειδικούς εκτός του χώρου, διασφαλίζοντας ότι οι ασθενείς λαμβάνουν έγκαιρες και ακριβείς διαγνώσεις. Επιπλέον, τα μοντέλα αυτά μπορούν να βοηθήσουν στην ταξινόμηση των περιπτώσεων, δίνοντας προτεραιότητα σε εκείνες που απαιτούν άμεση προσοχή και διευκολύνοντας την αποτελεσματικότερη χρήση των περιορισμένων πόρων υγειονομικής περίθαλψης [26, 11, 25].

Ε.8 Συμπέρασμα

Συνοπτικά, η συνεχής πρόοδος και ενσωμάτωση των μοντέλων λεζάντας ιατρικών εικόνων με βάση την τεχνητή νοημοσύνη υπόσχεται πολλά για τον μετασχηματισμό της παροχής υγειονομικής περίθαλψης. Βελτιώνοντας την ακρίβεια, την αποτελεσματικότητα και την προσβασιμότητα της ερμηνείας των ιατρικών εικόνων, τα μοντέλα αυτά μπορούν να συμβάλουν σε καλύτερα αποτελέσματα για τους ασθενείς και σε ένα αποτελεσματικότερο σύστημα υγειονομικής περίθαλψης.

Κεφάλαιο F

Συμπεράσματα

F.1 Σύνοψη των συνεισφορών

Στην παρούσα ερευνητική εργασία παρουσιάζεται μια ολοκληρωμένη προσέγγιση για τον υποτιτλισμό ιατρικών εικόνων με την ενσωμάτωση των νευρωνικών δικτύων συνελικτικής ανάλυσης (CNN), των νευρωνικών δικτύων επαναληπτικής ανάλυσης (RNN) και των μηχανισμών προσοχής. Οι βασικές συνεισφορές και τα ευρήματα αυτής της έρευνας είναι οι εξής:

- **Ανάπτυξη υβριδικού μοντέλου:**
 - **Αρχιτεκτονική του μοντέλου:** Ανάπτυξη ενός υβριδικού μοντέλου CNN-RNN ειδικά προσαρμοσμένο για τον υποτιτλισμό ιατρικών εικόνων. Το στοιχείο CNN υπερέχει στην εξαγωγή ιεραρχικών οπτικών χαρακτηριστικών από ιατρικές εικόνες, αποτυπώνοντας τόσο λεπτομέρειες χαμηλού επιπέδου όσο και σημασιολογία υψηλού επιπέδου. Το συστατικό RNN, ιδίως με αμφίδρομα στρώματα LSTM, παράγει συνεκτικές και σχετικές με το πλαίσιο λεζάντες με βάση αυτά τα εξαγόμενα χαρακτηριστικά [26, 11].
 - **Μηχανισμοί προσοχής:** Η ενσωμάτωση μηχανισμών προσοχής ενισχύει σημαντικά την απόδοση του μοντέλου, επιτρέποντάς του να εστιάζει στις πιο σχετικές περιοχές της εικόνας. Αυτή η επιλεκτική εστίαση βοηθάει αρκετά στην ιατρική απεικόνιση, όπου συγκεκριμένες περιοχές μπορεί να περιέχουν κρίσιμες διαγνωστικές πληροφορίες [26, 11, 6].
 - **Αμφίδρομα στρώματα LSTM:** Η ενσωμάτωση αμφίδρομων στρωμάτων LSTM

βελτιώνει την ικανότητα του μοντέλου να συλλαμβάνει εξαρτήσεις μεγάλης εμβέλειας και να κατανοεί τη διαδοχική φύση της γλώσσας, διασφαλίζοντας ότι οι παραγόμενες λεζάντες είναι τόσο ακριβείς όσο και κατάλληλες για το πλαίσιο [26, 9].

- **Ποσοτική απόδοση:**

- **Μετρήσεις αξιολόγησης:** Το μοντέλο πέτυχε ανταγωνιστικές μετρήσεις απόδοσης, με βαθμολογίες BLEU-1 και BLEU-2 0.309 και 0.205, αντίστοιχα. Αυτές οι βαθμολογίες υποδηλώνουν την επάρκεια του μοντέλου στη δημιουργία υποτίτλων [26].

- **Ποιοτική ανάλυση:**

- **Παραδείγματα λεζάντων:** Ποιοτικά παραδείγματα παραγόμενων λεζάντων ανέδειξαν την ικανότητα του μοντέλου να παράγει κλινικά σχετικές και κατατοπιστικές περιγραφές. Τα παραδείγματα αυτά περιλάμβαναν ακριβείς περιγραφές φυσιολογικών ευρημάτων, προσδιορισμό παθολογικών χαρακτηριστικών και παροχή κατάλληλων για το πλαίσιο ιατρικών πληροφοριών [26, 11].
- **Κλινική συνάφεια:** Οι ποιοτικές αξιολογήσεις υπογράμμισαν τη χρησιμότητα του μοντέλου στην ακριβή περιγραφή σύνθετων ιατρικών εικόνων, καθιστώντας το πολύτιμο εργαλείο για την υποβοήθηση των επαγγελματιών υγείας στη διάγνωση και την τεκμηρίωση [26, 9].

- **Συγκριτική ανάλυση:**

- **Σύγκριση επιδόσεων:** Το προτεινόμενο μοντέλο ξεπέρασε τις υπάρχουσες σύγχρονες μεθόδους όσον αφορά τις βαθμολογίες BLEU, επικυρώνοντας την αποτελεσματικότητα της υβριδικής αρχιτεκτονικής CNN-RNN σε συνδυασμό με μηχανισμούς προσοχής. Αυτή η σύγκριση αποδεικνύει την υπεροχή της προσέγγισής μας στη δημιουργία υψηλής ποιότητας λεζάντες ιατρικών εικόνων [26, 13, 11, 8].

- **Περισσότερες επιπτώσεις:**

- **Επιπτώσεις στην παροχή υγειονομικής περίθαλψης:** Η παρούσα έρευνα υπογραμμίζει τις μετασχηματιστικές δυνατότητες της λεζάντας ιατρικών εικόνων με τεχνητή νοημοσύνη για την ενίσχυση της παροχής υγειονομικής περίθαλψης. Η

- αυτοματοποιημένη παραγωγή περιγραφικών λεζάντων μπορεί να βελτιώσει σημαντικά τη διαγνωστική αποτελεσματικότητα, μειώνοντας τον χρόνο που απαιτείται για την ερμηνεία των εικόνων και επιτρέποντας στους επαγγελματίες του τομέα της υγειονομικής περίθαλψης να επικεντρωθούν σε πιο σύνθετες περιπτώσεις [26, 8].
- **Διευκόλυνση της επικοινωνίας:** Οι ακριβείς και συναφείς με το πλαίσιο λεζάντες μπορούν να διευκολύνουν την καλύτερη επικοινωνία μεταξύ των επαγγελματιών υγείας, των ασθενών και άλλων ενδιαφερομένων. Αυτή η βελτιωμένη επικοινωνία μπορεί να οδηγήσει σε πιο τεκμηριωμένες κλινικές αποφάσεις και καλύτερο συντονισμό της περίθαλψης [26, 9, 22].
 - **Υποστήριξη της ιατρικής εκπαίδευσης:** Οι λεζάντες που δημιουργούνται με τεχνητή νοημοσύνη μπορούν να χρησιμεύσουν ως πολύτιμα εκπαιδευτικά εργαλεία, παρέχοντας στους εκπαιδευόμενους δομημένες και κατατοπιστικές περιγραφές ιατρικών εικόνων. Αυτό μπορεί να ενισχύσει τη μάθηση και να βελτιώσει τις διαγνωστικές δεξιότητες των επαγγελματιών υγείας [26, 9].
 - **Βοήθεια στην έρευνα:** Οι αυτοματοποιημένες λεζάντες μπορούν να βοηθήσουν στην ανάλυση συνόλων δεδομένων ιατρικής απεικόνισης μεγάλης κλίμακας, επιταχύνοντας τις ερευνητικές προσπάθειες για τη διάγνωση ασθενειών, το σχεδιασμό θεραπείας και τη θεραπευτική παρακολούθηση. Με την παροχή δομημένων σχολίων, τα μοντέλα αυτά μπορούν να υποστηρίξουν την έρευνα με βάση τα δεδομένα και την ανάπτυξη νέων κλινικών γνώσεων [26, 17, 14].
 - **Βελτίωση της τηλεϊατρικής:** Σε απομακρυσμένες ή υποεξυπηρετούμενες περιοχές, τα μοντέλα υποτιτλισμού με βάση την τεχνητή νοημοσύνη μπορούν να υποστηρίξουν την τηλεϊατρική παρέχοντας ακριβείς και έγκαιρες ερμηνείες ιατρικών εικόνων. Αυτό μπορεί να βελτιώσει την πρόσβαση σε εξειδικευμένη περίθαλψη και να βελτιώσει την ποιότητα των απομακρυσμένων διαβουλεύσεων [26, 11, 25].

F.2 Τελικές σκέψεις

Η επιτυχής ενσωμάτωση των νευρωνικών δικτύων με συνελκτικό τρόπο (CNN) και των νευρωνικών δικτύων με επαναλαμβανόμενο τρόπο (RNN) στη λεζάντα ιατρικών εικόνων αποτελεί σημαντικό ορόσημο στην υγειονομική περίθαλψη με βάση την τεχνητή νοημοσύνη. Το υβριδικό αυτό μοντέλο, που συχνά συνδυάζεται με τεχνικές μεταφοράς μάθησης και

προσαρμογής μοντέλων, επιδεικνύει αξιοσημείωτη προσαρμοστικότητα και γενίκευση σε διάφορες απεικονιστικές μεθόδους και παθολογίες, επιτρέποντας τη δημιουργία περιγραφικών και κλινικά σχετικών λεζάντων που περικλείουν τις περιπλοκές των ιατρικών εικόνων. Τα CNNs αναγνωρίζουν και εξάγουν χαρακτηριστικά από τις εικόνες, ενώ τα RNNs χειρίζονται την αλληλουχία των δεδομένων για να συνθέσουν κείμενο που είναι συνεκτικό και σχετικό με το περιεχόμενο της εικόνας. Αυτή η προσέγγιση έχει αποδειχθεί αποτελεσματική σε διάφορους τομείς της ιατρικής απεικόνισης, συμπεριλαμβανομένων των ακτινογραφιών, των μαγνητικών τομογραφιών και των υπερήχων, παρέχοντας έναν ισχυρό εργαλείο για τους επαγγελματίες υγείας για την καλύτερη διάγνωση και θεραπεία ασθενειών [26, 11, 9, 6].

Περαιτέρω, η ενσωμάτωση τεχνικών προ-εκπαίδευσης και η χρήση μεγάλων δεδομένων (big data) συμβάλλει στη βελτίωση της ακρίβειας και της απόδοσης αυτών των μοντέλων. Τα νευρωνικά δίκτυα ενισχυμένα με προ-εκπαίδευση σε μεγάλες βάσεις δεδομένων ιατρικών εικόνων μπορούν να εντοπίσουν λεπτές ανωμαλίες και πρότυπα που μπορεί να διαφεύγουν από την ανθρώπινη παρατήρηση, καθιστώντας τα εξαιρετικά χρήσιμα σε κλινικά περιβάλλοντα με περιορισμένους πόρους. Επιπλέον, οι πρόσφατες εξελίξεις στην ερμηνευσιμότητα των μοντέλων AI, όπως οι τεχνικές ανάλυσης προσοχής και οι χάρτες θερμότητας, επιτρέπουν στους επαγγελματίες υγείας να κατανοήσουν καλύτερα τις αποφάσεις που λαμβάνονται από τα μοντέλα, αυξάνοντας την εμπιστοσύνη και την υιοθέτηση αυτών των τεχνολογιών.

Πέρα από τα διαγνωστικά: Ευρύτερες επιπτώσεις

• Ιατρική εκπαίδευση:

- **Εκπαιδευτικός πόρος:** Η λεζάντα ιατρικών εικόνων με βάση την τεχνητή νοημοσύνη παρέχει έναν ανεκτίμητο πόρο για την ιατρική εκπαίδευση. Οι αυτοματοποιημένοι σχολιασμοί μπορούν να χρησιμεύσουν ως διδακτικά βοηθήματα, προσφέροντας δομημένες και κατατοπιστικές περιγραφές που ενισχύουν τη μάθηση. Οι εκπαιδευόμενοι μπορούν να χρησιμοποιήσουν αυτές τις λεζάντες που δημιουργούνται από τεχνητή νοημοσύνη για να κατανοήσουν τις αποχρώσεις των διαφόρων ιατρικών καταστάσεων και των απεικονιστικών ευρημάτων, βελτιώνοντας τις διαγνωστικές τους δεξιότητες και τις κλινικές τους γνώσεις [26, 9].
- **Διαδραστική μάθηση:** Η τεχνολογία μπορεί επίσης να υποστηρίξει διαδραστικά εργαλεία μάθησης, όπως πλατφόρμες εικονικής πραγματικότητας (VR) ή επαυξημένης πραγματικότητας (AR), όπου οι μαθητές μπορούν να εμπλακούν με σχολιασμένες ιατρικές εικόνες σε ένα δυναμικό περιβάλλον, εμβαθύνοντας την κατανόησή

τους μέσω καθηλωτικών εμπειριών [26].

- **Έρευνα:**

- **Ανακάλυψη με βάση τα δεδομένα:** Η δυνατότητα δημιουργίας ακριβών λεζάντων για μεγάλα σύνολα δεδομένων ιατρικών εικόνων επιταχύνει τις ερευνητικές προσπάθειες. Οι δομημένοι σχολιασμοί διευκολύνουν την ανάλυση των δεδομένων απεικόνισης, επιτρέποντας στους ερευνητές να εντοπίζουν μοτίβα, συσχετίσεις και τάσεις που μπορεί να μην είναι εμφανείς μέσω της χειροκίνητης ανάλυσης. Αυτό μπορεί να οδηγήσει σε νέες γνώσεις στον χαρακτηρισμό ασθενειών, στην πρόβλεψη της ανταπόκρισης στη θεραπεία και σε άλλους κρίσιμους τομείς [26, 17, 14].
- **Συνεργατικές προσπάθειες:** Η λεζάντα εικόνας με τεχνητή νοημοσύνη προωθεί τη συνεργατική έρευνα παρέχοντας ένα κοινό πλαίσιο για την ερμηνεία και την ανταλλαγή ιατρικών εικόνων. Ερευνητές από διαφορετικά ιδρύματα μπορούν να συνεργάζονται πιο αποτελεσματικά, αξιοποιώντας κοινά σύνολα δεδομένων με σχολιασμό για την προώθηση της καινοτομίας και της ανακάλυψης [26].

- **Παροχή υγειονομικής περίθαλψης:**

- **Βελτιωμένη λήψη κλινικών αποφάσεων:** Παρέχοντας λεπτομερείς και ακριβείς περιγραφές ιατρικών εικόνων, τα μοντέλα λεζάντας με βάση την τεχνητή νοημοσύνη υποστηρίζουν τους κλινικούς γιατρούς στη λήψη πιο τεκμηριωμένων και έγκαιρων αποφάσεων. Αυτό μπορεί να βελτιώσει τη διαγνωστική ακρίβεια, τον προγραμματισμό της θεραπείας και την παρακολούθηση του ασθενούς, οδηγώντας τελικά σε καλύτερα αποτελέσματα για τους ασθενείς [26, 11, 8].
- **Βελτιωμένη αποδοτικότητα:** Η αυτοματοποίηση της διαδικασίας δημιουργίας λεζάντας μειώνει τον φόρτο εργασίας των ακτινολόγων και άλλων επαγγελματιών υγείας, επιτρέποντάς τους να επικεντρωθούν σε πιο σύνθετες περιπτώσεις και εργασίες που απαιτούν την τεχνογνωσία τους. Αυτό μπορεί να ενισχύσει τη συνολική αποδοτικότητα της παροχής υγειονομικής περίθαλψης, μειώνοντας τα σημεία συμφόρησης και βελτιώνοντας την απόδοση των ασθενών [26].

F.3 Ενδυνάμωση των επαγγελματιών της υγειονομικής περίθαλψης

Η επιτυχής ενσωμάτωση των νευρωνικών δικτύων με συνελικτικό τρόπο (CNN) και των νευρωνικών δικτύων με επαναλαμβανόμενο τρόπο (RNN) στη λεζάντα ιατρικών εικόνων αποτελεί σημαντικό ορόσημο στην υγειονομική περίθαλψη με βάση την τεχνητή νοημοσύνη. Το υβριδικό αυτό μοντέλο, που συχνά συνδυάζεται με τεχνικές μεταφοράς μάθησης και προσαρμογής μοντέλων, επιδεικνύει αξιοσημείωτη προσαρμοστικότητα και γενίκευση σε διάφορες απεικονιστικές μεθόδους και παθολογίες, επιτρέποντας τη δημιουργία περιγραφικών και κλινικά σχετικών λεζάντων που περιλαμβάνουν τις περιπλοκές των ιατρικών εικόνων. Τα CNNs αναγνωρίζουν και εξάγουν χαρακτηριστικά από τις εικόνες, ενώ τα RNNs χειρίζονται την αλληλουχία των δεδομένων για να συνθέσουν κείμενο που είναι συνεκτικό και σχετικό με το περιεχόμενο της εικόνας. Αυτή η προσέγγιση έχει αποδειχθεί αποτελεσματική σε διάφορους τομείς της ιατρικής απεικόνισης, συμπεριλαμβανομένων των ακτινογραφιών, των μαγνητικών τομογραφιών και των υπερήχων, παρέχοντας έναν ισχυρό εργαλείο για τους επαγγελματίες υγείας για την καλύτερη διάγνωση και θεραπεία ασθενειών [26, 11, 9, 6].

Ο απώτερος στόχος της έρευνας αυτής είναι να ενδυναμώσει τους επαγγελματίες του τομέα της υγειονομικής περίθαλψης με καινοτόμα εργαλεία για την ενισχυμένη ερμηνεία ιατρικών εικόνων, τη λήψη κλινικών αποφάσεων και τη φροντίδα των ασθενών. Παρέχοντας ακριβείς και κλινικά σχετικές λεζάντες, το υβριδικό μοντέλο CNN-RNN βοηθά τους επαγγελματίες υγείας στην ερμηνεία σύνθετων απεικονιστικών ευρημάτων, διασφαλίζοντας ότι οι κρίσιμες πληροφορίες μεταδίδονται με σαφήνεια και αποτελεσματικότητα [26, 11, 8].

F.4 Προώθηση της διεπιστημονικής συνεργασίας και της καινοτομίας

Η επίτευξη του πλήρους δυναμικού της Τεχνητής Νοημοσύνης στην ιατρική απεικόνιση απαιτεί την προώθηση της διεπιστημονικής συνεργασίας, την ανταλλαγή γνώσεων και τη συνεχή καινοτομία. Οι ερευνητές, οι κλινικοί ιατροί και οι εταίροι της βιομηχανίας πρέπει να συνεργαστούν για να:

- **Προώθηση της τεχνολογίας:** Να συνεχίσουν να βελτιώνουν και να επεκτείνουν τις δυνατότητες των μοντέλων που βασίζονται στην ΤΝ μέσω συνεργατικών προσπα-

θειών έρευνας και ανάπτυξης. Αυτό περιλαμβάνει τη διερεύνηση νέων αρχιτεκτονικών, την ενσωμάτωση πολυτροπικών δεδομένων και τη βελτίωση της ερμηνευσιμότητας και της διαφάνειας [26, 11, 8].

- **Εξασφάλιση ηθικής ανάπτυξης:** Αντιμετωπίστε ηθικά ζητήματα που σχετίζονται με το απόρρητο των ασθενών, την ασφάλεια των δεδομένων και την αλγοριθμική προκατάληψη για να διασφαλίσετε ότι οι τεχνολογίες TN αναπτύσσονται υπεύθυνα και δίκαια. Ο καθορισμός ηθικών κατευθυντήριων γραμμών και ρυθμιστικών προτύπων είναι ζωτικής σημασίας για τη διατήρηση της εμπιστοσύνης και της αποδοχής μεταξύ των επαγγελματιών υγείας και των ασθενών [26].
- **Προώθηση της ανταλλαγής γνώσεων:** Διευκόλυνση της διάδοσης των ερευνητικών ευρημάτων και των βέλτιστων πρακτικών μέσω συνεδρίων, δημοσιεύσεων και συνεργατικών πλατφορμών. Αυτό θα βοηθήσει να διασφαλιστεί ότι οι εξελίξεις στον τομέα της λεζάντας ιατρικών εικόνων με βάση την τεχνητή νοημοσύνη θα είναι προσβάσιμες στην ευρύτερη κοινότητα της υγειονομικής περίθαλψης [26].
- **Ενθάρρυνση της συνεχούς μάθησης:** Υποστήριξη της συνεχούς εκπαίδευσης και κατάρτισης των επαγγελματιών του τομέα της υγειονομικής περίθαλψης, ώστε να τους βοηθήσει να κατανοήσουν και να χρησιμοποιήσουν αποτελεσματικά τα εργαλεία που βασίζονται στην TN. Αυτό περιλαμβάνει την παροχή πόρων και ευκαιριών για μάθηση σχετικά με τις τελευταίες εξελίξεις στην TN και τον τρόπο με τον οποίο μπορούν να εφαρμοστούν στην κλινική πρακτική [26].

F.5 Μετασχηματιστικές Αλλαγές στην Υγειονομική Περίθαλψη

Αξιοποιώντας τη δύναμη της Τεχνητής Νοημοσύνης, οι ερευνητές και οι επαγγελματίες υγείας μπορούν να οδηγήσουν σε μετασχηματιστικές αλλαγές στην υγειονομική περίθαλψη. Η λεζάντα ιατρικών εικόνων με βάση την τεχνητή νοημοσύνη έχει τη δυνατότητα να:

- **Βελτίωση των αποτελεσμάτων των ασθενών:** Βελτίωση της ακρίβειας και της αποτελεσματικότητας της διάγνωσης και του σχεδιασμού της θεραπείας, οδηγώντας σε καλύτερη φροντίδα και αποτελέσματα των ασθενών [26, 8].

- **Αύξηση της πρόσβασης στην περίθαλψη:** Υποστήριξη της τηλεϊατρικής και των εξ αποστάσεως διαβουλεύσεων, παρέχοντας έγκαιρες και ακριβείς ερμηνείες ιατρικών εικόνων σε υποβαθμισμένες περιοχές [26, 11, 25].
- **Βελτιστοποίηση της χρήσης των πόρων:** Βελτιστοποίηση των κλινικών ροών εργασίας και μείωση της επιβάρυνσης των επαγγελματιών υγείας, επιτρέποντας την αποτελεσματικότερη χρήση των πόρων υγείας [26].

Η επιτυχής ενσωμάτωση των νευρωνικών δικτύων με συνελικτικό τρόπο (CNN) και των νευρωνικών δικτύων με επαναλαμβανόμενο τρόπο (RNN) στη λεζάντα ιατρικών εικόνων αποτελεί σημαντικό ορόσημο στην υγειονομική περίθαλψη με βάση την τεχνητή νοημοσύνη. Το υβριδικό αυτό μοντέλο, που συχνά συνδυάζεται με τεχνικές μεταφοράς μάθησης και προσαρμογής μοντέλων, επιδεικνύει αξιοσημείωτη προσαρμοστικότητα και γενίκευση σε διάφορες απεικονιστικές μεθόδους και παθολογίες, επιτρέποντας τη δημιουργία περιγραφικών και κλινικά σχετικών λεζάντων που περιλαμβάνουν τις περιπλοκές των ιατρικών εικόνων. Τα CNNs αναγνωρίζουν και εξάγουν χαρακτηριστικά από τις εικόνες, ενώ τα RNNs χειρίζονται την αλληλουχία των δεδομένων για να συνθέσουν κείμενο που είναι συνεκτικό και σχετικό με το περιεχόμενο της εικόνας. Αυτή η προσέγγιση έχει αποδειχθεί αποτελεσματική σε διάφορους τομείς της ιατρικής απεικόνισης, συμπεριλαμβανομένων των ακτινογραφιών, των μαγνητικών τομογραφιών και των υπερήχων, παρέχοντας έναν ισχυρό εργαλείο για τους επαγγελματίες υγείας για την καλύτερη διάγνωση και θεραπεία ασθενειών.

Εν κατακλείδι, η ενσωμάτωση της TN στη λεζάντα των ιατρικών εικόνων αποτελεί σημαντική πρόοδο στην τεχνολογία της υγειονομικής περίθαλψης. Συνεχίζοντας να καινοτομεί και να συνεργάζεται μεταξύ των κλάδων, η κοινότητα της υγειονομικής περίθαλψης μπορεί να ξεκλειδώσει το πλήρες δυναμικό της TN, βελτιώνοντας τη φροντίδα των ασθενών, προωθώντας την έρευνα και ενισχύοντας την ιατρική εκπαίδευση. Αυτή η έρευνα αποτελεί ένα βήμα προς ένα μέλλον όπου τα εργαλεία που βασίζονται στην TN θα διαδραματίζουν κεντρικό ρόλο στην υγειονομική περίθαλψη, συμβάλλοντας τελικά στη βελτίωση των αποτελεσμάτων των ασθενών και στην υψηλότερη ποιότητα της περίθαλψης [26, 11, 8].

Βιβλιογραφία

- [1] Hareem Ayesha et al. “Automatic medical image interpretation: State of the art and future directions”. In: *Pattern Recognition* 114 (2021), p. 107856.
- [2] Jun Chen et al. “Visualgpt: Data-efficient adaptation of pretrained language models for image captioning”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 18030–18040.
- [3] Xinlei Chen and C Lawrence Zitnick. “Mind’s eye: A recurrent visual representation for image caption generation”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 2422–2431.
- [4] Murk Chohan et al. “Image captioning using deep learning: A systematic”. In: *image* 11.5 (2020).
- [5] Noel Codella et al. “Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic)”. In: *arXiv preprint arXiv:1902.03368* (2019).
- [6] Kelei He et al. “Transformers in medical image analysis”. In: *Intelligent Medicine* 3.1 (2023), pp. 59–78.
- [7] Michael Heath et al. “Current status of the digital database for screening mammography”. In: *Digital Mammography: Nijmegen, 1998*. Springer, 1998, pp. 457–460.
- [8] Benjamin Hou et al. “Ratchet: Medical transformer for chest x-ray diagnosis and reporting”. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VII 24*. Springer. 2021, pp. 293–303.

- [9] Jia-Hong Huang, Ting-Wei Wu, and Marcel Worring. “Contextualized keyword representations for multi-modal retinal image captioning”. In: *Proceedings of the 2021 International Conference on Multimedia Retrieval*. 2021, pp. 645–652.
- [10] Jia-Hong Huang et al. “Longer Version for” Deep Context-Encoding Network for Retinal Image Captioning”. In: *arXiv preprint arXiv:2105.14538* (2021).
- [11] Hojun Lee et al. “Cross encoder-decoder transformer with global-local visual extractor for medical image captioning”. In: *Sensors* 22.4 (2022), p. 1429.
- [12] Minh-Thang Luong, Hieu Pham, and Christopher D Manning. “Effective approaches to attention-based neural machine translation”. In: *arXiv preprint arXiv:1508.04025* (2015).
- [13] David Lyndon, Ashnil Kumar, Jinman Kim, et al. “Neural Captioning for the ImageCLEF 2017 Medical Image Challenges.” In: *CLEF (working notes)*. 2017.
- [14] Hyeryun Park et al. “Feature difference makes sense: a medical image captioning model exploiting feature difference and tag information”. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*. 2020, pp. 95–102.
- [15] Hyeryun Park et al. “Medical image captioning model to convey more details: Methodological comparison of feature difference generation”. In: *IEEE Access* 9 (2021), pp. 150560–150568.
- [16] Alexander Selivanov et al. “Medical image captioning via generative pretrained transformers”. In: *Scientific Reports* 13.1 (2023), p. 4171.
- [17] Dilbag Singh et al. “Efficient evolving deep ensemble medical image captioning network”. In: *IEEE Journal of Biomedical and Health Informatics* 27.2 (2022), pp. 1016–1025.
- [18] Sanjay Subramanian et al. “Medicat: A dataset of medical images, captions, and textual references”. In: *arXiv preprint arXiv:2010.06000* (2020).
- [19] Daniel Vela Jarquin. “Caption generation with transformer models across multiple medical imaging modalities”. In: ().
- [20] Oriol Vinyals et al. “Show and tell: A neural image caption generator”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 3156–3164.

- [21] Zhanyu Wang et al. “A medical semantic-assisted transformer for radiographic report generation”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2022, pp. 655–664.
- [22] Ting-Wei Wu et al. “Expert-defined keywords improve interpretability of retinal image captioning”. In: *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. 2023, pp. 1859–1868.
- [23] Mengya Xu, Mobarakol Islam, and Hongliang Ren. “Rethinking surgical captioning: End-to-end window-based mlp transformer using patches”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2022, pp. 376–386.
- [24] Di You et al. “Aligntransformer: Hierarchical alignment of visual regions and disease tags for medical report generation”. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24*. Springer. 2021, pp. 72–82.
- [25] Jinglu Zhang et al. “Surgical instruction generation with transformers”. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part IV 24*. Springer. 2021, pp. 290–299.
- [26] Wenshuo Zhou et al. “Transferring pre-trained large language-image model for medical image captioning”. In: *CLEF2023 Working Notes, CEUR Workshop Proceedings, CEUR-WS. org, Thessaloniki, Greece*. 2023.

Συντμήσεις

AI	Artificial Intelligence
CNN	Convolutional Neural Network
CT	Computed Tomography
CV	Computer Vision
DDSM	Digital Database for Screening Mammography
DEXA	Dual-Energy X-ray Absorptiometry
DTI	Diffusion Tensor Imaging
fMRI	Functional Magnetic Resonance Imaging
Faster R-CNN	Faster Region-based Convolutional Neural Network
GRU	Gated Recurrent Unit
ISIC	International Skin Imaging Collaboration

LDCT	Low-Dose Computed Tomography
LSTM	Long Short-Term Memory
MRI	Magnetic Resonance Imaging
OLIVER	Optimized Linguistic Interpretation and Visual Explanation Reporter
ReLU	Rectified Linear Unit
RNN	Recurrent Neural Network
U-NET	U-shaped Network
YOLO	You Only Look Once

Γλωσσάρι Ξενικών Όρων

Dot-product

Εσωτερικό Γινόμενο

Token

Δείγμα

Bi-gram

Δίγραμμα

Κώδικας

Libraries

```
1 import os
2 import pickle
3 import numpy as np
4
5 import tensorflow as tf
6 from tensorflow.keras.layers import Input
7 from tensorflow.keras.layers import Embedding
8 from tensorflow.keras.layers import LSTM
9 from tensorflow.keras.layers import Dense
10 from tensorflow.keras.layers import Dropout
11 from tensorflow.keras.layers import Add
12 from tensorflow.keras.layers import Bidirectional
13 from tensorflow.keras.layers import Concatenate
14 from tensorflow.keras.layers import LayerNormalization
15 from tensorflow.keras.models import Model
16 from tensorflow.keras.optimizers import Adam
17 from tensorflow.keras.callbacks import EarlyStopping
18 from tensorflow.keras.callbacks import ReduceLROnPlateau
19 from tensorflow.keras.callbacks import ModelCheckpoint
20 from tensorflow.keras.callbacks import TensorBoard
21 from tensorflow.keras.preprocessing.sequence import pad_sequences
22 from tensorflow.keras.utils import to_categorical
```

Data Loading

```
1 with open('/kaggle/input/dataset/image_features.pkl','rb') as a:
2     features=pickle.load(a)
3
4 with open('/kaggle/input/dataset/captions.pkl','rb') as b:
```

```
5 mapping=pickle.load(b)
6
7 with open('/kaggle/input/dataset/max_length.pkl','rb') as c:
8     max_length=pickle.load(c)
```

Text Tokenization

```
1 # Add start and end tokens to the captions
2 for image_id, captions in mapping.items():
3     for idx in range(len(captions)):
4         captions[idx] = f'<start> {captions[idx]} <end>'
5
6 # Collect all captions in a single list
7 all_captions = [caption for captions in mapping.values() for caption in ↵
8     captions]
9
10 # Tokenize the captions
11 tokenizer = Tokenizer()
12
13 tokenizer.fit_on_texts(all_captions)
14
15 # Calculate vocabulary size
16 vocab_size = len(tokenizer.word_index) + 1
```

Train/Test Split

```
1 image_ids=list(mapping.keys())
2 split=int(len(image_ids)*0.8)
3 train=image_ids[:split]
4 test=image_ids[split:]
```

Model Building

```
1 import tensorflow as tf
2
3 from tensorflow.keras.layers import Input
4 from tensorflow.keras.layers import Embedding
5 from tensorflow.keras.layers import LSTM
6 from tensorflow.keras.layers import Dense
7 from tensorflow.keras.layers import Dropout
8 from tensorflow.keras.layers import Bidirectional
9 from tensorflow.keras.layers import Add
10 from tensorflow.keras.layers import Attention
```

```

11 from tensorflow.keras.layers import Concatenate
12 from tensorflow.keras.layers import GlobalAveragePooling1D
13 from tensorflow.keras.models import Model
14 from tensorflow.keras.callbacks import EarlyStopping
15 from tensorflow.keras.callbacks import ReduceLROnPlateau
16 from tensorflow.keras.callbacks import ModelCheckpoint
17 from tensorflow.keras.callbacks import TensorBoard
18 from tensorflow.keras.optimizers import Adam
19 from tensorflow.keras.metrics import Precision
20 from tensorflow.keras.metrics import Recall
21
22
23 # Encoder
24 # Image feature model
25 inputs1 = Input(shape=(4096,))
26 fe1 = Dropout(0.5)(inputs1) # Adjusted dropout rate
27 fe2 = Dense(512, activation='relu')(fe1)
28
29 # Sequence feature layer
30 inputs2 = Input(shape=(max_length,))
31 se1 = Embedding(vocab_size, 256, mask_zero=True)(inputs2)
32 se2 = Dropout(0.5)(se1) # Adjusted dropout rate
33 # Using a Bidirectional LSTM
34 se3 = Bidirectional(LSTM(256, return_sequences=True))(se2)
35
36 # Attention mechanism
37 attn_out = Attention()([se3, se3])
38 attn_out = GlobalAveragePooling1D()(attn_out) # Use global average ↔
        pooling to get a fixed-size output
39
40 # Decoder
41 decoder1 = Add()([fe2, attn_out]) # Combine the image features and the ↔
        pooled attention output
42 decoder2 = Dense(256, activation='relu')(decoder1)
43 decoder2 = Dropout(0.5)(decoder2) # Add dropout for regularization
44 outputs = Dense(vocab_size, activation='softmax')(decoder2)
45
46 # Model definition
47 model = Model(inputs=[inputs1, inputs2], outputs=outputs)
48
49 # Learning rate schedule

```

```

50 lr_schedule = tf.keras.optimizers.schedules.ExponentialDecay(
51     initial_learning_rate=1e-4, # Adjusted for a more conservative start
52     decay_steps=10000,
53     decay_rate=0.9)
54
55 # Optimizer
56 optimizer = Adam(learning_rate=lr_schedule)
57
58 # Compile the model with the new optimizer and learning rate schedule
59 model.compile(optimizer=optimizer,
60               loss='categorical_crossentropy',
61               metrics=['accuracy', Precision(), Recall()])
62
63 # Print model summary to verify the architecture
64 model.summary()
65
66 # Callbacks for learning rate reduction, early stopping, model ↵
    checkpointing, and TensorBoard
67 reduce_lr = ReduceLROnPlateau(monitor='val_loss', factor=0.2, patience=5, ↵
    min_lr=1e-5)
68 checkpoint = ModelCheckpoint(filepath='/kaggle/working/model_best.h5', ↵
    save_best_only=True, monitor='val_loss', mode='min')
69 early_stop = EarlyStopping(monitor='val_loss', patience=10, mode='min')
70 tensorboard_callback = TensorBoard(log_dir='/kaggle/working/logs', ↵
    histogram_freq=1)

```

Model Training

```

1 import os
2 from tqdm import tqdm
3
4 import tensorflow as tf
5 from tensorflow.keras.models import load_model
6 from tensorflow.keras.optimizers import Adam
7 from tensorflow.keras.metrics import Precision
8 from tensorflow.keras.metrics import Recall
9 from tensorflow.keras.callbacks import ModelCheckpoint
10 from tensorflow.keras.callbacks import EarlyStopping
11 from tensorflow.keras.callbacks import ReduceLROnPlateau
12 from tensorflow.keras.callbacks import TensorBoard
13
14

```

[illegible]

```

multiprocessing if supported
49         workers=2, # Adjust based on your system's capabilities
50         max_queue_size=10)
51
52     # Save model every 10 epochs
53     if (epoch + 1) % 10 == 0:
54         model_save_path = f'/kaggle/working/model_epoch_{epoch + 1}.h5'
55         model.save(model_save_path)
56         print(f"Model saved to {model_save_path}")
57
58     except Exception as e:
59         print(f"Error during training in epoch {epoch + 1}: {e}")
60         continue
61
62 # Save the final model
63 model.save('/kaggle/working/model_final.h5')
64 print("Final model saved to /kaggle/working/model_final.h5")

```

Caption Generation

```

1 def create_index_to_word_dict(tokenizer):
2     # Invert the tokenizer.word_index to create a new dictionary
3     index_word_dict = {index: word for word, index in tokenizer.
4     word_index.items()}
5     return index_word_dict
6
7 def index_to_word(integer, index_word_dict):
8     # Access the word directly using the integer as a key
9     return index_word_dict.get(integer, None)
10
11 import tensorflow as tf
12 from tensorflow.keras.models import load_model
13 # Loading the model
14 model_path = '/kaggle/input/cnn-rnn-att/model_epoch_50.h5'
15 model = load_model(model_path)
16
17 def predict_caption(model, image, tokenizer, max_length):
18     in_text = '<start>'
19     for _ in range(max_length):
20         sequence = tokenizer.texts_to_sequences([in_text])[0]

```

```

20         sequence = pad_sequences([sequence], maxlen=max_length)
21         yhat = model.predict([image, sequence], verbose=0)
22         yhat = np.argmax(yhat)
23         word = tokenizer.index_word[yhat]
24         if word is None:
25             break
26         in_text += ' ' + word
27         if word == '<end>':
28             break
29     return in_text
30
31 from tensorflow.keras.preprocessing.sequence import pad_sequences
32 import numpy as np
33
34 def predict_caption(model, image, tokenizer, max_length):
35     in_text = '<start>'
36     for _ in range(max_length):
37         # Encode the input text as a sequence of integers
38         sequence = tokenizer.texts_to_sequences([in_text])[0]
39         # Pad the sequence to ensure it has the same length as max_length
40         sequence = pad_sequences([sequence], maxlen=max_length)
41         # Predict the next word in the sequence
42         yhat = model.predict([image, sequence], verbose=0)
43         yhat = np.argmax(yhat)
44         # Convert the integer to the corresponding word
45         word = tokenizer.index_word.get(yhat)
46         if word is None:
47             break
48         in_text += ' ' + word
49         if word == '<end>':
50             break
51     # Remove the start and end tokens from the generated caption
52     caption = in_text.replace('<start> ', '').replace(' <end>', '')
53     return caption

```

Model Evaluation

```

1 from nltk.translate.bleu_score import corpus_bleu
2 from tqdm import tqdm
3
4 def evaluate_model(model, test_keys, mapping, features, tokenizer, ↵
    max_length):

```

```

5     actual, predicted = list(), list()
6     for key in tqdm(test_keys):
7         if key in mapping and key in features:
8             # Preprocess features and captions
9             feature = features[key]
10            captions = mapping[key]
11            actual_captions = [caption.split() for caption in captions]
12
13            # Generate prediction
14            y_pred = predict_caption(model, feature, tokenizer, ←
max_length)
15            y_pred_words = y_pred.split()
16
17            # Collect actual and predicted captions
18            actual.append(actual_captions)
19            predicted.append(y_pred_words)
20
21            # Calculate BLEU scores
22            bleu_scores = {
23                'BLEU-1': corpus_bleu(actual, predicted, weights=(1.0, 0, 0, 0)),
24                'BLEU-2': corpus_bleu(actual, predicted, weights=(0.5, 0.5, 0, 0)←
),
25                'BLEU-3': corpus_bleu(actual, predicted, weights=(0.3, 0.3, 0.3, ←
0)),
26                'BLEU-4': corpus_bleu(actual, predicted, weights=(0.25, 0.25, ←
0.25, 0.25))
27            }
28
29            return bleu_scores
30
31 # Usage example
32 test_keys = list(mapping.keys())
33 bleu_scores = evaluate_model(model, test_keys, mapping, features, ←
tokenizer, max_length)
34
35 # Save evaluation results to a text file
36 file_path = '/kaggle/working/evaluation_results.txt'
37 with open(file_path, 'w') as f:
38     f.write(f"BLEU Scores: {bleu_scores}")
39
40 # Print evaluation results

```

41 `print("BLEU Scores:", bleu_scores)`

Πίνακας F.1: Model Summary: "OLIVER"

Layer (type)	Output Shape	Param #	Connected to
input_4 (InputLayer)	(None, 22)	0	[]
embedding_1 (Embedding)	(None, 22, 256)	6045440	['input_4[0][0]']
dropout_3 (Dropout)	(None, 22, 256)	0	['embedding_1[0][0]']
input_3 (InputLayer)	(None, 4096)	0	[]
bidirectional_1 (Bidirectional)	(None, 22, 512)	1050624	['dropout_3[0][0]']
dropout_2 (Dropout)	(None, 4096)	0	['input_3[0][0]']
attention (Attention)	(None, 22, 512)	0	['bidirectional_1[0][0]', 'bidirectional_1[0][0]']
dense_3 (Dense)	(None, 512)	2097664	['dropout_2[0][0]']
global_average_pooling1d (GlobalAveragePooling1D)	(None, 512)	0	['attention[0][0]']
add_1 (Add)	(None, 512)	0	['dense_3[0][0]', 'global_average_pooling1d[0][0]']
dense_4 (Dense)	(None, 256)	131328	['add_1[0][0]']
dropout_4 (Dropout)	(None, 256)	0	['dense_4[0][0]']
dense_5 (Dense)	(None, 23615)	6069055	['dropout_4[0][0]']
Total params:	15394111 (58.72 MB)		
Trainable params:	15394111 (58.72 MB)		
Non-trainable params:	0 (0.00 Byte)		