دانشکده مهندسی کامپیوتر و فناوری اطلاعات

پیشنهاد موضوع پایان نامه (Proposal)

تولید خودکار گزارش تصاویر رادیولوژی با استفاده از شبکه های عصبی عمیق

Automatic generation of radiology image reports using deep neural network

دانشجو

اسماء صنعتی (۴۰۱۲۱۴۰۱۶)

کارشناسی ارشد مهندسی کامپیوتر – شبکه های کامپیوتری

استاد راهنما

دکتر رسول اسماعیلی فرد

استاد مشاور

دکتر -----

۱. مقدمه(Introduction)

زمینه پژوهش

تصویربرداری پزشکی یکی از اجزای حیاتی مراقبت‌های بهداشتی مدرن است که اطلاعات ضروری برای تشخیص، درمان و نظارت بر بیماری‌های مختلف را فراهم می‌کند. یادگیری عمیق، به عنوان یکی از زیرشاخه‌های یادگیری ماشین، در سال‌های اخیر تحولات چشمگیری در زمینه تصویربرداری پزشکی ایجاد کرده است. این تکنولوژی به واسطه توسعه شبکه‌های عصبی مصنوعی با لایه‌های مخفی متعدد، قادر به یادگیری خودکار و استخراج ویژگی‌های سلسله مراتبی از داده‌های خام است. برخلاف روش‌های سنتی یادگیری ماشین، مدل‌های یادگیری عمیق توانایی بالاتری در تحلیل و تفسیر داده‌های پیچیده دارند [۱].

توموگرافی کامپیوتری با پرتو مخروطی (CBCT) یکی از فناوری‌های پیشرفته تصویربرداری پزشکی است که قابلیت تولید تصاویر سه‌بعدی دقیق از بدن را دارد [۲]. این تصاویر اغلب به همراه گزارش‌هایی که توسط رادیولوژیست‌ها تهیه می‌شوند، ارائه می‌گردند. تاکنون پیشرفت‌های قابل توجهی در زمینه پیاده‌سازی مدل‌های تولید خودکار گزارش‌های رادیولوژی به دست آمده است، که این پیشرفت‌ها عمدتاً به کاربردهای موفق یادگیری عمیق در این حوزه نسبت داده می‌شوند [۳]. تحقیقات فعلی بر روی ترکیب پردازش زبان طبیعی و بینایی کامپیوتری برای بهبود دقت و کارایی گزارش‌های خودکار متمرکز هستند . [۴]

اهمیت و ضرورت موضوع

هدف اصلی این تحقیق بررسی ساختارهای فک و صورت در تصاویر رادیولوژی و ایجاد ارتباط بین این ساختارها و گزارش‌های رادیولوژیست‌ها با استفاده از یادگیری عمیق است. این تحقیق دو مزیت عمده را دنبال می‌کند: افزایش دقت در تشخیص بیماری‌ها با بهره‌گیری از قابلیت‌های هوش مصنوعی، و تسریع فرآیند ارائه گزارش‌های رادیولوژی به کمک تولید خودکار گزارش‌ها. این مزایا می‌توانند به بهبود کیفیت مراقبت‌های بهداشتی و کاهش بار کاری رادیولوژیست‌ها منجر شوند [۵].

نوآوری و هدف

این پژوهش بر روی تولید خودکار گزارش‌های رادیولوژی برای تصاویر CBCT مفصل گیجگاهی (TMJ) متمرکز است، که یکی از مفاصل حیاتی در ناحیه فک و صورت می‌باشد. با توجه به محدودیت مطالعات پیشین در این زمینه، این پژوهش اولین تلاش برای تولید خودکار گزارش‌های رادیولوژی از تصاویر CBCT مفصل گیجگاهی است. هدف از این مطالعه ارائه یک راهکار با دقت بالا برای تشخیص و گزارش‌دهی بیماری‌های مرتبط با این مفصل است . [۶]

استفاده از مدل‌های یادگیری عمیق مانند VGG-16 و LSTM در این تحقیق، امکان پردازش دقیق و تولید گزارش‌های متنی جامع و دقیق را فراهم می‌کند. این ترکیب تکنیک‌ها می‌تواند تحولی مهم در بهبود فرآیندهای تشخیصی و گزارش‌دهی در حوزه رادیولوژی ایجاد کند[7] .

۲. بیان مسأله(Problem Statement)

شرح دقیق مسأله

مطالعه حاضر بر روی توموگرافی کامپیوتری با پرتو مخروطی (CBCT) تمرکز دارد و از تصاویر CBCT بدست آمده از مفصل گیجگاهی (TMJ) که یکی از مفاصل حیاتی در ناحیه فک و صورت محسوب می‌شود، استفاده خواهد کرد تا بتواند گزارش رادیولوژی مرتبط با بیماری‌های مفصل گیجگاهی را بصورت خودکار تولید نماید. این مفصل نقش مهمی در عملکرد فک دارد و مشکلات مرتبط با آن می‌تواند منجر به درد و محدودیت در حرکت فک شود. تشخیص دقیق و سریع این مشکلات برای درمان مؤثر ضروری است.

چالش‌ها و مشکلات موجود

با توجه به اهمیت تشخیص دقیق و سریع بیماری‌های مفصل گیجگاهی، تولید خودکار گزارش‌های رادیولوژی از تصاویر CBCT این ناحیه، چالشی مهم در حوزه هوش مصنوعی و تصویربرداری پزشکی به شمار می‌رود. تاکنون، مطالعات انجام‌گرفته برای تولید گزارش‌های رادیولوژی از روی تصاویر رادیولوژی فک و صورت بسیار محدود بوده است و این پژوهش اولین مطالعه در زمینه تولید خودکار گزارش‌های رادیولوژی بر روی تصاویر CBCT از مفصل گیجگاهی (TMJ) است [۱]. ارائه یک راهکار با خطای پایین برای تشخیص بیماری‌های مرتبط با این مفصل از اهمیت بالایی برخوردار است..

۳. سوالات پژوهش(Research Questions)

در این پایان‌نامه به سوالات زیر پرداخته خواهد شد:

- آیا می‌توان یک سیستم خودکار برای تولید گزارش رادیولوژی از تصاویر CBCT مفصل گیجگاهی (TMJ) ایجاد نمود؟
- معیارهای ارزیابی مرتبط با دقت و خطای این سیستم چگونه خواهند بود؟
- چه تکنیک‌ها و مدل‌های یادگیری عمیق برای بهبود دقت و کارایی این سیستم مناسب هستند؟
- چگونه می‌توان اطمینان حاصل کرد که گزارش‌های تولید شده توسط سیستم خودکار با گزارش‌های تهیه شده توسط رادیولوژیست‌ها همخوانی دارند و از دقت بالایی برخوردارند؟

این سوالات به دنبال یافتن راه‌حل‌هایی برای بهبود فرآیند تشخیص و گزارش‌دهی بیماری‌های مفصل گیجگاهی با استفاده از تکنیک‌های پیشرفته یادگیری عمیق و پردازش تصاویر پزشکی هستند. هدف اصلی این پژوهش ارائه یک راهکار کارآمد و دقیق برای تسریع فرآیندهای تشخیصی و کاهش خطاهای انسانی در گزارش‌دهی رادیولوژی است.

۴. پیشینه پژوهش(Literature Review)

مرور مقالات و پژوهش‌های مرتبط

مروری بر تکنیک‌های پردازش تصاویر پزشکی

جدول ۱ تعدادی از مدل‌های مورد استفاده در پردازش تصاویر پزشکی با استفاده از یادگیری عمیق مبتنی بر Convolutional Neural Network (CNN) را که تا کنون در تصویربرداری پزشکی و وظایف تشخیصی استفاده می‌شوند، ارائه کرده است. این مدل‌ها شامل معماری‌های متنوعی مانند AlexNet، VGG-16، VGG-19، GoogLeNet، ResNet-101، DenseNet و ResNet-152 هستند. تمرکز بر کاربرد این مدل‌ها در طیف وسیعی از وظایف تصویربرداری پزشکی است، مانند تشخیص بیماری‌های متعدد (مانند آدنوپاتی، متاستاز، بیماری‌های سینوسی، پنومونی، آمفیزم، برونشیت) و در بخش‌های مختلف بدن (مانند قفسه سینه، گردن، استخوان، کبد، مغز، قلب). دیتاست‌های ذکر شده شامل PACS مرکز بالینی NIH، PACS of the fourth people's hospital ، ChestX-ray14 و IU X-Ray هستند. روش‌های تصویربرداری مختلفی از جمله CT، MR، PET، سونوگرافی و X-ray نیز پوشش داده شده‌اند.

جدول ۱. تکنیک‌های مختلف مبتنی بر CNN در تشخیص تصویربرداری پزشکی[۳]

مدل	پیشنهاد شده توسط	حالت تصویر	دیتاست	عضو	آسیب شناسی	نرم افزار	معماری CNN	تکنیک
Deep mining model	Shin 2015 [7]	CT MR PET	PACS of NIH clinical centre	چندگانه (به عنوان مثال گردن،استخوان، کبد، مغز و قلب)	چندگانه (به عنوان مثال آدنوپاتی، متاستاز و بیماریهای سینوسی)	caffe	AlexNet VGG-16 VGG-19	LDA & RNN CNN
LDPO: looped deep pseudo task optimization network	Wang, et al. 2016 [8]	سونوگرافی رادیوگرافی کامپیوتری				Caffe	AlexNet GoogLeNet	CNN K-means/RIM NLP PCA

CNN-based classification model	Dong, et al. 2017 [9]	X-Ray	PACS of the fourth people's hospital (Chinees reports)	قفسه سینه	۹ بیماری (مانند آمفیزم و برونشیت)	Caffe	VGG-16 ResNet-101	NLP CNN RNN
CheXNet	Rajpurkar, 2017 [۱۰]	ChestX-ray14		قفسه سینه	پنومونی و ۱۳ آسیب شناسی دیگر	-	DenseNet	CNN CAM
ChestNet	Wang and Xia 2018 [11]	ChestX-ray14		قفسه سینه		Caffe	Resnet-152	CNN Attention mechanism (Grad-CAM)
DualNet	Rubin, et al. 2019[12]	ChestX-ray14	MIMIC-CXR	قفسه سینه	۱۴ بیماری قفسه سینه (مانند ذات الریه)	PyTorch	DenseNet-121	NLP CNN
Multi-task learning model	Jing, et al. 2017 [13]	X-Ray	IU X-Ray	قفسه سینه	بیماری های قفسه سینه	-	VGG-19	CNN hierarchical LSTM MLC
Multimodal recurrent model with attention	Xue, et al. 2018 [14]	X-Ray	IU X-Ray	قفسه سینه	بیماری های قفسه سینه	-	Resnet-152	CNN Single layer LSTM Bi-LSTM and ID CNN
TieNet: text-image embedding network	Wang, et al. 2018 [15]	X-Ray	IU X-Ray	قفسه سینه	بیماری های قفسه سینه	TensorFlow Tensorpack	ResNet-50	NLP CNN-RNN LSTM-RNN
-	Chang et al., T2Re 2022[16]	CT	پایگاه داده تحقیقاتی کلینیکی دانشگاه پزشکی تایپه	ریه	سرطان ریه	Keras	-	NLP, CNN, RNN
ISCM-LBP + PCA-HOG, XceptionNet, Modified-MHA	Dhruv Sharma, Chhavi Dhiman, Dinesh Kumar (2024)[17]	X-Ray	IU Chest X-ray	قفسه سینه	چندگانه	-	C-TaXNet, Dr2T (Dense Radiology Report Generation Transformer	FDT-Dr2T

تکنیک های مورد استفاده برای تولید خودکار گزارش

جدول ۲ تعدادی از مدل‌هایی را که برای تولید گزارش‌های ساختاریافته استفاده شده اند نمایش داده است. در آموزش این مدل ها اغلب از مجموعه داده ImageNet [۱۸] استفاده شده است. این مدل‌ها از معماری‌هایی مانند VGG-16، ResNet-50، FCDenseNet و DeepSPINE بهره می‌برند. تکنیک‌های مورد استفاده در آنها شامل استخراج و پیش‌بینی FFL (یافته‌ها، ویژگی‌ها و ضایعات) با استفاده از معماری‌های CNN، طبقه‌بندی چند برچسبی و تولید گزارش مبتنی بر درخت تصمیم هستند. این مدل‌ها بر تولید گزارش‌های ساختاریافته برای روش‌های تصویربرداری خاص مانند X-ray (قفسه سینه)، MRI (ستون فقرات کمری) و CT (بخش‌های مختلف بدن مانند کبد) تمرکز داشته اند.

جدول ۲. مدل‌های پیش‌آموزش دیده برای تولید گزارش‌های ساختاریافته در تصویربرداری پزشکی[۱۸]

منبع	حوزه کاری	تکنیک مورد استفاده	مسئله مورد توجه	مدل هوش مصنوعی مورد استفاده
CNN-FFL (2020) [19]	X-ray: chest	-یافتن برچسب های مناسب(FFL) -استخراج FFL از گزارش ها و ساخت dataset -پیش‌بینی FFL با استفاده از معماری CNN به‌ویژه از یک FPN که ترکیبی از VGG-16 و ResNet-50 است. -استفاده از بلوک‌های dilated در طراحی شبکه	تولید گزارش ساختار یافته	Pre-trained VGG-16 and ResNet-50 on ImageNet
MultusRadBot (2020)[۲۰]	MRI: lumbar spine	-استفاده از (CNN)، به ویژه FCDenseNet و DeepSPINE به همراه تجزیه و تحلیل اجزای اصلی (PCA). -استفاده از دو CNN، مانند VGG ، برای کار طبقه‌بندی چند کلاسه -استفاده از درخت تصمیم برای تولید گزارش	طبقه بندی چند برچسبی در نهایت: کامپایل به گزارش ها	ندارد
Pre-LesaNet (2019) [21]	CT: 115 body parts, 27 types, and 29 attributes	CNN (VGG-16)	تولید خودکار گزارش‌های دقیق و ساختاریافته مرتبط با مجموعه‌ای از برچسب‌ها	**ندارد**
CNN-SVM (2021) [22]	CT: liver	-استخراج ویژگی‌های تصویر با استفاده از CNN، به‌ویژه MobileNet و LeNet-5 -استفاده از SVM با یک هسته خطی به عنوان طبقه‌بندی کننده	تولید گزارش ساختار یافته	**MobileNet that trained on ImageNet**
CNN-RNN (2020) [23]	X-ray: chest	استفاده از معماری CNN GoogLeNet برای نمایش تصویر و متعاقبا استفاده از LSTM یا GRU برای تولید متن	تولید گزارش ساختار یافته	**ندارد**

Sequence CNN-RNN [24]	X-ray: chest	-استفاده از معماری CNN VGG16 /ResNet-50 برای نمایش تصویر -استفاده از LSTM برای تولید متن	تولید گزارش ساختاریافته	Pre-trained VGG-16 /ResNet-50 that trained on ImageNet

مقایسه تکنیک ها

در هر دو جدول، از مدل یادگیری عمیق مبتنی بر CNN استفاده شده است. جدول اول طیف گسترده‌ای از معماری‌ها مانند AlexNet، سری VGG، سری ResNet، DenseNet و GoogLeNet را برجسته می‌کند که در فریم‌ورک‌های نرم‌افزاری مختلف از جمله Caffe، PyTorch و TensorFlow به کار می‌روند. در مقابل، جدول دوم بر مدل‌هایی تمرکز دارد که در ImageNet آموزش دیده‌اند و از معماری‌هایی مانند VGG-16، ResNet-50، MobileNet و LeNet-5 بهره می‌برند. فریم‌ورک‌های نرم‌افزاری ذکر شده در جدول دوم به وظایفی مانند تولید گزارش ساختاریافته اختصاص دارند و شامل TensorFlow، Tensorpack و ترکیباتی از CNN و RNN (LSTM یا GRU) هستند.

کاربردها و بیماری‌ها

کاربردهای متنوعی در جدول ۱ مورد بررسی قرار گرفته است و از شناسایی چندین بیماری در بخش‌های مختلف بدن تا بیماری‌های خاص مانند پنومونی و بیماری‌های قفسه سینه را شامل می‌شوند. این جدول نشان می دهد که مجموعه داده هایی مانند ChestX-ray14 و IU X-Ray بصورت گسترده در تشخیصی مورد استفاده قرار گرفته است. در مقابل، جدول ۲ بیشتر بر تولید گزارش‌های ساختاریافته پزشکی تمرکز دارد و شرایط خاصی مانند اختلالات ستون فقرات کمری و شرایط کبد تا کنون بیشتر مورد توجه بوده. این جدول همچنین دیتاست‌هایی مرتبط با این وظایف را نشان می‌دهد.

روش‌های تصویربرداری

مطابق با جدول ۱ روش‌های تصویربرداری متنوعی شامل CT، MR، PET، سونوگرافی و X-ray تا کنون مورد توجه قرار گرفته اند. این تنوع نشان می دهد که مدل‌های مبتنی بر CNN در تشخیص تصویربرداری پزشکی موفق بوده اند. با این حال، جدول دوم نشان می دهد که اغلب مطالعات بر روی موضوعاتی مانند X-ray (قفسه سینه)، MRI (ستون فقرات کمری) و CT (کبد) تمرکز داشته اند که نشان‌دهنده رویکرد هدفمندتری در تولید گزارش های رادیولوژی است.

نوآوری‌ها

مدل‌های موجود در جدول اول اغلب ترکیبات و تکنیک‌های جدیدی مانند ترکیب CNN با RNN (به عنوان مثال، LDA و RNN)، یا استفاده از مکانیزم‌های attention (به عنوان مثال، CheXNet)، و رویکردهای

یادگیری چند وظیفه‌ای ارائه کرده اند. در این نوآوری‌ها هدف بهبود دقت تشخیصی و تفسیر تصاویر پزشکی بوده است. جدول دوم بر نوآوری‌ها در تولید گزارش را به تصویر کشیده است که اغلب آنها بر روی فرمت‌های خروجی ساختاریافته و طبقه‌بندی‌های چند برچسبی، بهبود قابلیت اطمینان و جامعیت گزارش‌های پزشکی خودکار تمرکز داشته اند.

رویه استفاده شده در مطالعات قبلی مطابق جدول ۱ و ۲ برای تولید خودکار گزارش های رادیولوژی به شرح زیر بوده است:

۱. **استخراج ویژگی‌های بصری با استفاده از CNN:** در این روش های ذکر شده در جدول اول ابتدا از روش هایی مانندAlexNet ، VGG-16، ResNet-101و DenseNet ، برای استخراج ویژگی‌های مهم و اطلاعات بصری از تصاویر پزشکی استفاده شده است. این ویژگی‌های بصری شامل اطلاعات مهمی مانند ناهنجاری‌ها، توده‌ها و سایر ویژگی‌های مرتبط با بیماری‌ها می‌باشند.
۲. **به‌کارگیری مکانیزم‌هایAttention:** برخی از مدل‌ها مانند CheXNet از مکانیزم‌های توجه برای تمرکز بر بخش‌های مهم تصویر استفاده می‌کنند. این مکانیزم‌ها به مدل کمک می‌کنند تا بر روی نواحی مهم تصویر که اطلاعات بیشتری دارند، تمرکز کند و گزارش‌های دقیق‌تری تولید کند.
۳. **تبدیل ویژگی‌های بصری به متن با استفاده از RNN :**پس از استخراج ویژگی‌های بصری، این اطلاعات به مدل‌های زبانی مانند LSTM یا GRU داده می‌شود که وظیفه تولید گزارش‌های متنی را دارند. به عنوان مثال، از ترکیباتی مانندCNN و RNN LSTMیا GRU برای تولید گزارش‌های دقیق و جامع استفاده می‌شود.

مقایسه و تحلیل پژوهش‌های قبلی با پژوهش پیشنهادی

این پژوهش در زمینه تولید خودکار گزارش های رادیولوژی بر روی تصاویر CBCT از مفصل گیجگاهی (TMJ) است که در مقایسه با پژوهش های قبلی بررسی تصاویر CBCT یک نوآوری است.

۵. روش تحقیق(Research Methodology)

توضیح روش پیشنهادی

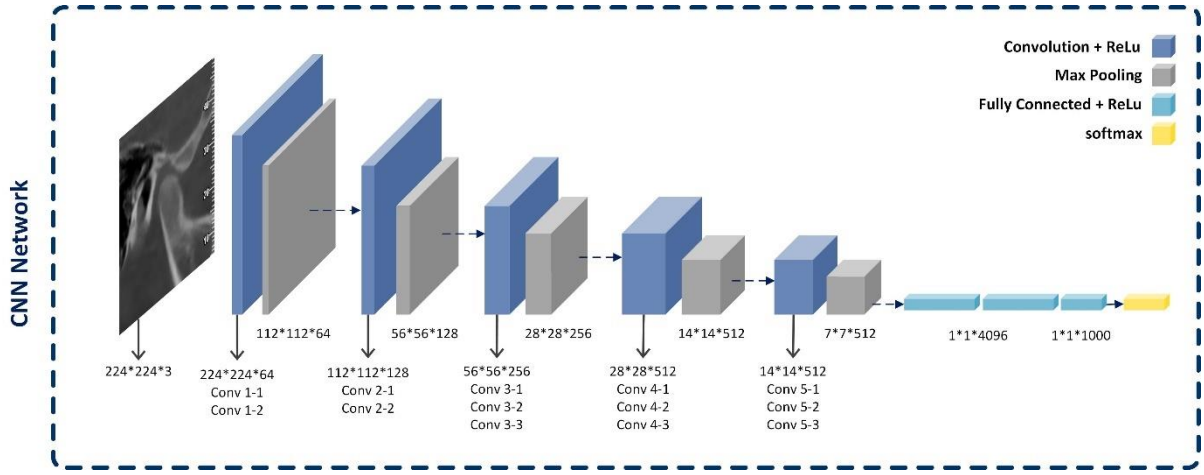
پردازش تصاویر بر مبنای شبکه عصبی عمیق VGG

در این راهکار برای پردازش تصاویر، از یک شبکه عصبی عمیق به نام۱۶-VGG استفاده خواهد شد و تصاویر CBCT مفصل گیجگاهی (TMJ) به عنوان ورودی به آن ارائه خواهد شد.

VGG-۱۶ یکی از مهم‌ترین و تأثیرگذارترین معماری‌های شبکه عصبی کانولوشنی است که در سال ۲۰۱۴ در دانشگاه آکسفورد معرفی شد [۲۵]. این معماری نوآورانه در زمان خود یک پیشرفت قابل توجه در زمینه بینایی ماشین و یادگیری عمیق به شمار می‌رفت. این شبکه با هدف بررسی تأثیر عمق شبکه بر دقت در وظایف تشخیص تصویر در مقیاس بزرگ، VGG-۱۶ طراحی شد. VGG-۱۶به سرعت به یکی از پرکاربردترین مدل‌های پیش‌آموزش دیده در جامعه یادگیری ماشین تبدیل شد.

محققان و مهندسان از این مدل برای انواع مختلفی از وظایف بینایی ماشین، از جمله تشخیص اشیاء، بخش‌بندی تصویر و حتی در زمینه‌های خاصی مانند تشخیص پزشکی استفاده کردند. قابلیت انتقال‌پذیری بالای VGG-۱۶ که به توانایی آن در یادگیری ویژگی‌های عمومی و قابل استفاده در وظایف مختلف اشاره دارد، آن را به یک انتخاب محبوب برای یادگیری انتقالی تبدیل کرد. این می تواند تصاویر CBCT را پردازش کرده و ویژگی‌های مهم و اساسی تصاویر را استخراج نماید.

VGG-۱۶ یک شبکه پیش‌آموزش داده‌شده روی دیتاست ImageNet است. دیتاست ImageNet شامل بیش از ۱۴ میلیون تصویر است که به صورت دستی برچسب‌گذاری شده‌اند. این تصاویر به ۱۰۰۰ کلاس مختلف تقسیم شده‌اند که طیف گسترده‌ای از اشیاء و موجودات را پوشش می‌دهند. این مدل با دقت Top-۵ حدود ۹۲.۷٪ در مجموعه داده ImageNet نشان داده است که افزایش عمق شبکه می‌تواند به بهبود قابل توجهی در دقت منجر شود. عدد ۱۶ در معماری VGG-۱۶ به تعداد کل لایه‌ها با پارامترهای قابل آموزش (لایه‌های پیچشی و لایه‌های کاملاً متصل) اشاره دارد. شکل ۱ معماری مورد استفاده از این شبکه را نشان می دهد.



شکل ۱. معماری مدل VGG پیشنهادی

- **۱۳ لایه Convolutional**: این لایه‌ها مسئول استخراج ویژگی‌ها از تصاویر ورودی هستند. آنها از فیلترهای کوچک با اندازه ۳×۳ پیکسل استفاده می‌کنند که این یک از ویژگی منحصربفرد VGG-۱۶ است. این انتخاب کمک می‌کند تا جزئیات دقیق در تصاویر را بدست آورند. لایه‌های Convolutional به پنج بلوک تقسیم می‌شوند که هر کدام توسط یک لایه max-pooling دنبال می‌شوند.
- **۳ لایه Fully Connected (Dense)**: پس از لایه‌های Convolutional وmax pooling ، سه لایه کاملاً متصل (FC) وجود دارد. دو لایه اول FC هر کدام ۴۰۹۶ نورون دارند و لایه سوم دارای ۱۰۰۰ نورون است که به تعداد کلاس‌های موجود در مجموعه داده ImageNet مربوط می‌شود.
- **لایه‌های Max-Pooling**: پنج لایه Max-Pooling بین بلوک‌های Convolutional پراکنده شده‌اند. هر لایه pooling ابعاد فضایی نقشه‌های ویژگی را به نصف کاهش می‌دهد که این کار به کاهش بار محاسباتی و کنترل برفرازگیری کمک می‌کند.
- **تابع فعال‌سازی:** تمام لایه‌های مخفی از تابع فعال‌سازی ReLU استفاده می‌کنند ReLU .غیرخطی بودن مدل را معرفی می‌کند و امکان یادگیری الگوهای پیچیده‌تر را فراهم می‌کند.
- **لایه Softmax:** لایه نهایی VGG-۱۶ یک تابع فعال‌ساز softmax است که توزیع احتمالات بر روی ۱۰۰۰ کلاس خروجی را تولید می‌کند.

آموزش شبکه‌های عصبی عمیق مانند VGG-۱۶ چالش‌های خاص خود را دارد. برای غلبه بر این چالش‌ها و بهبود عملکرد مدل، محققان از چندین تکنیک پیشرفته آموزشی استفاده کردند که هر کدام نقش مهمی در موفقیت نهایی مدل داشتند. یکی از مهم‌ترین تکنیک‌های مورد استفاده، Data Augmentationیا افزایش داده است. این روش به طور قابل توجهی تنوع داده‌های آموزشی را افزایش می‌دهد، بدون نیاز به جمع‌آوری داده‌های بیشتر. تکنیک‌های افزایش داده شامل چرخش تصادفی تصاویر، تغییر مقیاس، برش تصادفی، و تغییرات جزئی در روشنایی بود. این تکنیک به مدل کمک می‌کند تا ویژگی‌های مقاوم‌تر و عمومی‌تری را یاد بگیرد که منجر به عملکرد بهتر در تصاویر جدید و ناشناخته می‌شود. علاوه بر این، افزایش داده به کاهش over-fittingکمک می‌کند، زیرا مدل مجبور است الگوهای کلی‌تری را یاد بگیرد که در انواع مختلف نمایش یک تصویر صادق باشند.

تکنیک دیگری که در آموزش VGG-۱۶ مورد استفاده قرار گرفت،Dropout است. این روش یک راهکار ساده اما بسیار مؤثر برای جلوگیری از over-fitting است. در طول آموزش، dropout به طور تصادفی تعدادی از نورون‌ها را در هر تکرار غیرفعال می‌کند. این کار باعث می‌شود که شبکه نتواند بیش از حد به ویژگی‌های خاصی وابسته شود و در عوض، یادگیری توزیع‌شده‌تر و مقاوم‌تری داشته باشد. درVGG-۱۶ ، dropout عمدتاً

در لایه‌های کاملاً متصل استفاده شده است، جایی که خطر over-fitting به دلیل تعداد زیاد پارامترها بیشتر است. استفاده از dropout باعث شد که مدل بتواند عملکرد بهتری در داده‌های جدید داشته باشد و قابلیت تعمیم‌پذیری آن افزایش یابد.

گرچه در نسخه اصلی VGG-۱۶ از Batch Normalization استفاده نشد، این تکنیک در نسخه‌های بعدی و در بسیاری از کاربردهای مدرن VGG-۱۶ مورد استفاده قرار گرفته است Batch Normalization .یک تکنیک قدرتمند است که به تسریع همگرایی در طول آموزش کمک می‌کند. این روش با نرمال‌سازی خروجی هر لایه قبل از ورود به لایه بعدی، مشکل covariate shift را کاهش می‌دهد. به این ترتیب، هر لایه می‌تواند یادگیری را با سرعت بیشتری انجام دهد، زیرا نیازی به سازگاری مداوم با تغییرات توزیع ورودی ندارد. علاوه بر تسریع آموزش، Batch Normalization به ثبات بیشتر در طول آموزش کمک می‌کند و حتی می‌تواند به عنوان یک نوع منظم‌سازی عمل کند، که باز هم به جلوگیری از over-fitting کمک می‌کند.

استفاده از این تکنیک‌ها در کنار هم، نقش مهمی در موفقیت VGG-۱۶ و مدل‌های مشابه داشته است. آنها نه تنها به بهبود دقت مدل کمک کردند، بلکه آموزش مدل‌های عمیق‌تر را امکان‌پذیر ساختند.

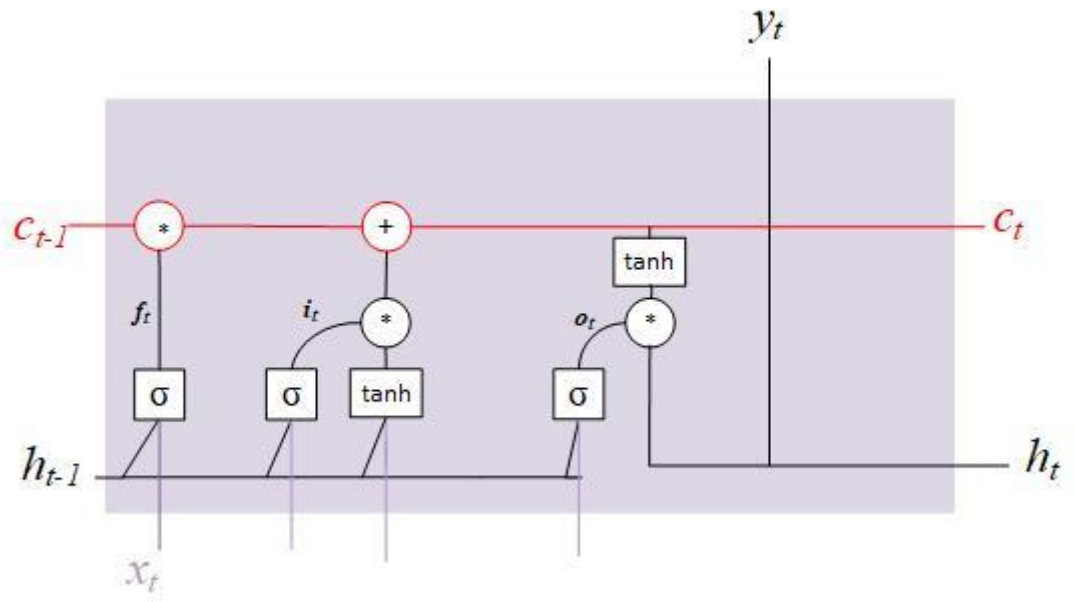
تبدیل ویژگی های بصری به گزارش با آموزش مدل زبانی LSTM

پس از استخراج و شناسایی ویژگی های تصویر و با استفاده از یک شبکه LSTM تولید گزارش خودکار انجام می گیرد. گزارش خودکار حاوی توصیفی از جزئیات و نتایج مهم موجود در تصاویر خواهد بود. بدین منظور، در کنار هر یک از تصاویر یک فایل متنی که حاوی گزارش آن تصویر قرار می گیرد. بدین منظور از یک مدل شبکه حافظه کوتاه مدت بلند مدت (LSTM) به عنوان مدل زبانی برای تولید گزارش های متنی استفاده خواهد شد. LSTM قادر است در هر مرحله از ورودی، کلمه بعدی را پیش‌بینی کرده و بر اساس ویژگی‌های تصویری و دنباله‌های ورودی شرح، یک گزارش متنی مرتبط و جامع تولید کند.

شبکه‌های حافظه‌ی بلند کوتاه‌مدت (LSTM) در زمینه یادگیری عمیق نقش مهمی ایفا می‌کنند و به‌ویژه در حوزه‌هایی مانند پردازش زبان طبیعی و دنباله‌های زمانی کاربردهای گسترده‌ای دارند. این شبکه‌ها از ساختار پیچیده‌ای بهره می‌برند که امکان حفظ وابستگی‌های طولانی‌مدت در داده‌ها را فراهم می‌آورد. به عنوان مثال، در مسائلی مانند ترجمه ماشینی، تولید شرح تصاویر، یا تحلیل داده‌های زمانی، LSTM می‌توانند اطلاعات طولانی‌مدت را در دنباله‌ها تجزیه و تحلیل کنند که باعث بهبود دقت و کارایی مدل می‌شود.

حافظه کوتاه‌مدت بلند (LSTM) یک نوع خاص از شبکه عصبی بازگشتی (RNN) است که برای پردازش و پیش‌بینی داده‌های ترتیبی طراحی شده است. این شبکه برای رفع مشکلات شبکه‌های عصبی بازگشتی معمولی، به ویژه مشکل ناپدید شدن گرادیان، توسعه یافته است [۲۶] و از آن زمان تاکنون به عنوان یکی از

قدرتمندترین ابزارهای یادگیری عمیق در پردازش داده‌های ترتیبی شناخته شده اند. ساختار LSTM از سلول‌های حافظه و مکانیزم‌های دروازه‌ای تشکیل شده است که به آن‌ها امکان می‌دهد اطلاعات را برای مدت‌های طولانی‌تر حفظ کرده و به یاد بیاورند. این مکانیزم‌ها شامل دروازه ورودی، دروازه فراموشی و دروازه خروجی هستند. معماری این بلاک این شبکه در شکل ۳ نمایش داده شده است.



شکل۲. معماری بلاک های LSTM

این مدل از وضعیت سلول قبلی h_{t-1}، بردار x_t و بایاس b به عنوان ورودی استفاده می کند و محتوای حافظه c_t را ایجاد می کند و h_t را به عنوان خروجی با انجام عملیاتی روی محتوای حافظه c_{t-1} قبلی تولید می کند. همچنین، چهار دروازه ذکر شده در بالا محتوای سلول حافظه را تنظیم می کنند. دروازه فراموشی عددی بین ۰ و ۱ تولید می کند تا مشخص کند که مقدار سلول حافظه قبلی تا چه میزان باید نادیده گرفته شود. مقدار نزدیک به صفر به این معنی است که بیشتر محتوای حافظه قبلی باید در فعلی فراموش می‌شود، در حالی که مقدار نزدیک معکوس آن را نمایش می دهد. عملیات دروازه فراموشی به شرح زیر است:

$$f_t = \sigma_g(w_f x_t + u_f h_{t-1} + b_f) \qquad (1)$$

گیت ورودی مشخص می کند که چه میزان از ورودی فعلی باید به حافظه اضافه شود. این مقدار بصورت تابعی بصورت زیر عمل می کند:

$$i_t = \sigma_g(w_i x_t + u_i h_{t-1} + b_i) \qquad (۲)$$

که در اینجا σ_g نشان دهنده تابع سیگموئید و w و u هم وزن هایی هستند که به منظور جلوگیری از مشکل از بین رفتن گرادیان استفاده می شوند. گیت کنترل بروز رسانی محتوای سلول حافظه از c_{t-1} به c_t بر اساس خروجی های دروازه ورودی و فراموشی نظارت می کند.

$$c_t = f_t . c_{t-1} + i_t \sigma_g(w_c x_t + u_c h_{t-1} + b_c) \qquad (۳)$$

دروازه خروجی نتیجه نهایی را تولید می کند و وضعیت سلول را از h_{t-1} به h_t تغییر می دهد. عملکرد این دروازه توسط توابع زیر تعریف می شود:

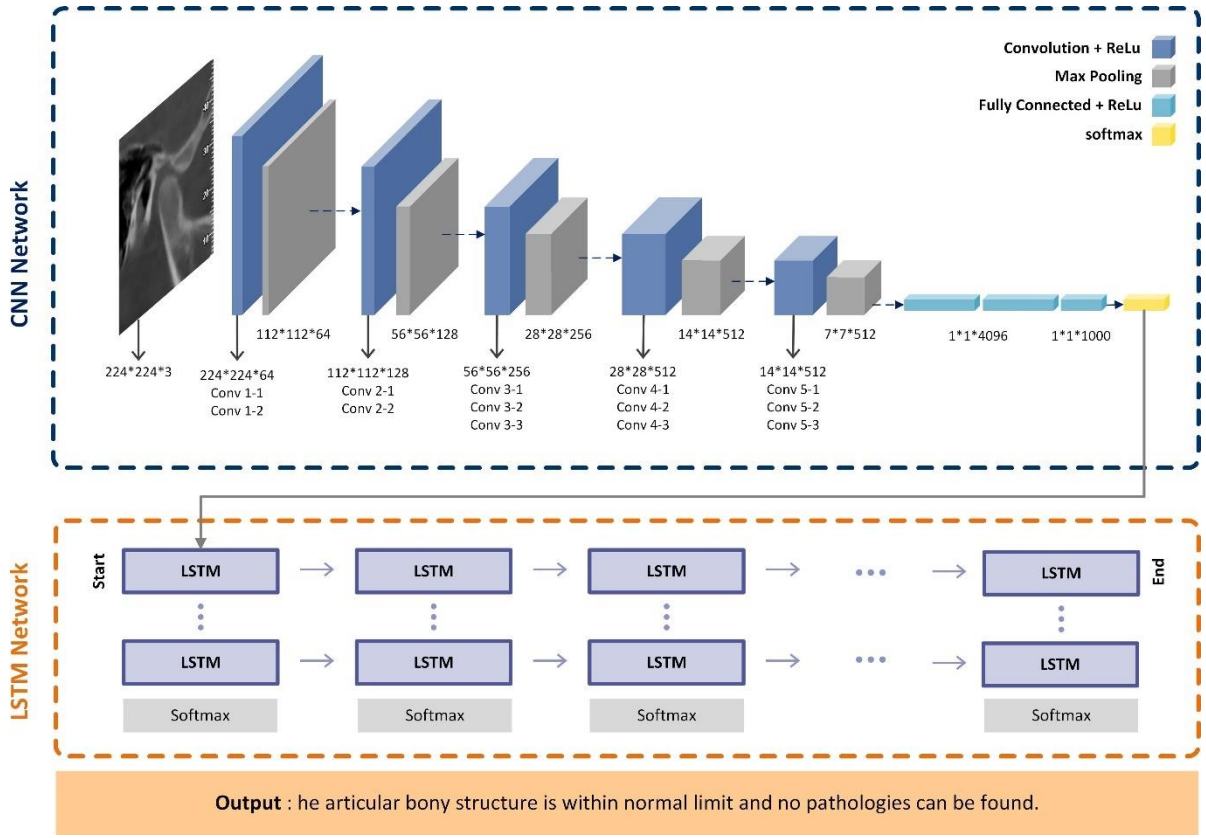
$$o_t = \sigma_g(w_o x_t + u_o h_{t-1} + b_o) \qquad (۴)$$

$$h_t = o_t . tanh(c_t) \qquad (۵)$$

که در اینجا tanh نشان دهنده تابع تانژانت هایپلبولیک است. توضیحات بیشتر در خصوص این مدل و مکانیزم انتشار رو به جلو و انتشار رو به عقب (backpropagation) در [۲۷] ارائه شده است.

معماری روش پیشنهادی

روش پیشنهادی بر معماری ترکیبی از شبکه عصبی کانولوشنی (CNN) و حافظه بلند مدت کوتاه (LSTM) خواهد بود. شکل ۳ تصویری از این معماری را نشان می‌دهد که برای تحلیل تصاویر و تولید گزارش متنی استفاده می‌شود. در این مدل، ابتدا تصویر ورودی توسط شبکه CNN پردازش می‌شود تا ویژگی‌های آن استخراج شود و سپس این ویژگی‌ها به شبکه LSTM داده می‌شود تا متن توصیفی گزارش را تولید کند.



شکل ۳. معماری روش پیشنهادی

در مرحله اول، تصویر ورودی که یک تصویر CBCT از مفصل گیجگاهی (TMJ) با اندازه۳*۲۲۴*۲۲۴ است، به مدل داده می‌شود. تصویر ابتدا از چندین لایه کانولوشن با فیلترهای ۳×۳ عبور می‌کند که هر کدام به یک تابع فعال‌سازی ReLU (Rectified Linear Unit) اعمال می‌شوند. این لایه‌ها ویژگی‌های تصویری مانند لبه‌ها، زوایا و بافت‌ها را استخراج می‌کنند. اندازه خروجی این لایه‌ها به تدریج کاهش می‌یابد و تعداد فیلترها افزایش می‌یابد.

پس از هر چند لایه کانولوشنی، یک لایه max-pooling قرار دارد که ابعاد فضایی (عرض و ارتفاع) خروجی را به نصف کاهش می‌دهد. این کار به کاهش بار محاسباتی و افزایش کارایی مدل کمک می‌کند. بعد از عبور از لایه‌های کانولوشن و max-pooling، خروجی به چندین لایه کاملاً متصل (Dense) داده می‌شود. این لایه‌ها شامل ۴۰۹۶ نورون در هر لایه هستند و اطلاعات تصویری به یک بردار ویژگی فشرده تبدیل می‌شود. در نهایت، یک لایه Softmax با ۱۰۰۰ نورون برای تولید احتمال تعلق تصویر به هر یک از کلاس‌های موجود در مجموعه داده ImageNet استفاده می‌شود.

در مرحله دوم، خروجی بردار ویژگی فشرده از شبکه VGG-16 به عنوان ورودی به شبکه LSTM داده می‌شود. شبکه LSTM شامل چندین لایه LSTM است که به ترتیب زمانی پردازش می‌شوند. این شبکه توانایی یادگیری و به خاطر سپردن توالی‌ها و وابستگی‌های طولانی‌مدت در داده‌ها را دارد. شبکه LSTM با گرفتن بردار ویژگی‌ها و پردازش آن‌ها، توالی کلمات متناظر با ویژگی‌های تصویری را تولید می‌کند. هر گام زمانی در LSTM با یک لایه Softmax دنبال می‌شود که توزیع احتمال برای انتخاب کلمه بعدی را فراهم می‌کند.

در نهایت، مدل یک متن توصیفی تولید می‌کند که محتوای تصویر را شرح می‌دهد. به عنوان مثال، در این نمودار، مدل گزارشی مانند " The articular bony structure is within normal limit and no pathologies can be found." را تولید می‌کند.

معیارهای ارزیابی

معیار BLEU

معیار BLEU (BiLingual Evaluation Understudy) برای ارزیابی کیفیت شرح‌های تولید شده توسط مدل استفاده می‌شود. این معیار میزان انطباق بین شرح‌های تولید شده توسط مدل و شرح‌های مرجع را اندازه‌گیری می‌کند BLEU .یک معیار خودکار است که کیفیت ترجمه‌های ماشینی را با مقایسه آن‌ها با چندین ترجمه مرجع انسانی ارزیابی می‌کند.

نحوه محاسبه :BLEU

n-گرام:

BLEUاز n-گرام‌ها (n-grams) برای مقایسه استفاده می‌کند n-گرام‌ها توالی‌هایی از n کلمه متوالی در متن هستند. برای مثال، در BLEU-1 از تک‌کلمه‌ای‌ها(unigrams) ، در BLEU-2 از دوکلمه‌ای‌ها(bigrams) ، در BLEU-3از سه‌کلمه‌ای‌ها (trigrams) و در BLEU-4 از چهارکلمه‌ای‌ها (quadgrams) استفاده می‌شود.

محاسبه Precision

دقت n- (Precision) گرام‌ها در شرح‌های تولید شده با n-گرام‌های مشابه در شرح‌های مرجع مقایسه می‌شود.

محاسبه Brevity Penalty (BP)

این معیار از جریمه کوتاهی (Brevity Penalty) برای جلوگیری از تولید جملات بسیار کوتاه که ممکن است دقت بالایی داشته باشند اما معنی‌دار نباشند، استفاده می‌کند.

محاسبه BLEU Score

امتیاز BLEU به صورت میانگین هندسی دقت-n گرام‌ها ضرب در جریمه کوتاهی محاسبه می‌شود.

معمولاً امتیاز BLEU بین ۰ و ۱ است که هرچه به ۱ نزدیک‌تر باشد، نشان‌دهنده انطباق بیشتر بین شرح‌های تولید شده و شرح‌های مرجع است.

انواع BLEU

BLEU-1: از تک‌کلمه‌ای‌ها (unigrams) استفاده می‌کند و دقت در سطح کلمات فردی را اندازه‌گیری می‌کند.

BLEU-2: از دوکلمه‌ای‌ها (bigrams) استفاده می‌کند و دقت در سطح ترکیبات دوکلمه‌ای را اندازه‌گیری می‌کند.

BLEU-3: از سه‌کلمه‌ای‌ها (trigrams) استفاده می‌کند و دقت در سطح ترکیبات سه‌کلمه‌ای را اندازه‌گیری می‌کند.

BLEU-4: از چهارکلمه‌ای‌ها (quadgrams) استفاده می‌کند و دقت در سطح ترکیبات چهارکلمه‌ای را اندازه‌گیری می‌کند.

معیار ROUGE (Recall-Oriented Understudy for Gisting Evaluation)

معیار ROUGE برای ارزیابی کیفیت خلاصه‌ها و شرح‌های تولید شده استفاده می‌شود. این معیار تمرکز بیشتری بر بازیابی اطلاعات (Recall) دارد و انواع مختلفی مانند ROUGE-N (n-گرام‌ها)، ROUGE-L (بلندترین زیرتوالی مشترک) و ROUGE-W (n-گرام‌های وزن‌دار) را شامل می‌شود.

معیار METEOR (Metric for Evaluation of Translation with Explicit ORdering)

METEOR برای ارزیابی کیفیت ترجمه‌های ماشینی استفاده می‌شود و به دقت (Precision) و بازیابی (Recall) توامان توجه دارد. این معیار همچنین از اطلاعات معنایی و ترکیب‌های زبانی استفاده می‌کند.

معیار CIDEr (Consensus-based Image Description Evaluation)

معیار CIDEr برای ارزیابی کیفیت توضیحات تصاویر تولید شده استفاده می‌شود. این معیار بر اساس توافق میان توضیحات تولید شده و توضیحات مرجع انسانی عمل می‌کند.

٦. روش پیاده‌سازی(Implementation Method)

زبان برنامه‌نویسی، محیط شبیه‌سازی و کتابخانه ها

رای پیاده‌سازی روش پیشنهادی تولید خودکار گزارش رادیولوژی با استفاده از VGG-16 و LSTM، از زبان برنامه‌نویسی پایتون و کتابخانه‌های زیر استفاده می‌شود:

TensorFlow: یک کتابخانه منبع باز برای محاسبات عددی و یادگیری ماشین که توسط تیم Google Brain توسعه یافته است. این کتابخانه امکان ساخت و آموزش مدل‌های یادگیری عمیق را با استفاده از گراف‌های محاسباتی فراهم می‌کند و از اجرای محاسبات روی CPU و GPU پشتیبانی می‌کند.

Keras: یک API سطح بالا برای ساخت و آموزش مدل‌های یادگیری عمیق که به دلیل سادگی و قابلیت استفاده آسان محبوبیت زیادی دارد. Keras امکان ساخت مدل‌های پیچیده با استفاده از چند خط کد را فراهم می‌کند و از شبکه‌های عصبی مختلفی مانند شبکه‌های عصبی پیچشی (CNN) و شبکه‌های عصبی بازگشتی (RNN) پشتیبانی می‌کند.

NumPy: برای محاسبات عددی.

Pandas: برای مدیریت و تجزیه و تحلیل داده‌ها.

Matplotlib: برای تجسم داده‌ها و نتایج.

شرح فرآیند پیاده‌سازی

مراحل پیاده‌سازی شامل مراحل زیر است:

پیش‌پردازش داده‌ها:

آماده‌سازی و پاکسازی داده‌ها برای اطمینان از کیفیت و انسجام آن‌ها.

استخراج ویژگی‌ها با استفاده از VGG16:

استفاده از مدل از پیش آموزش‌دیده VGG16 برای استخراج ویژگی‌های مهم از تصاویر رادیولوژی.

تولید متن با استفاده از LSTM:

استفاده از شبکه‌های عصبی بازگشتی LSTM برای تولید متن گزارش رادیولوژی بر اساس ویژگی‌های استخراج‌شده.

آموزش و ارزیابی مدل:

آموزش مدل با استفاده از داده‌های آموزشی و ارزیابی عملکرد آن با استفاده از داده‌های آزمون.

تنظیم مدل برای بهبود دقت و کارایی آن در تولید گزارش‌های دقیق رادیولوژی.

این فرآیندها به دقت و کارایی مدل در تولید گزارش‌های دقیق رادیولوژی کمک می‌کنند.

۷. داده‌های مورد استفاده (Data and Resources)

در این پژوهش، داده‌های مورد استفاده شامل تصاویر توموگرافی کامپیوتری با پرتو مخروطی (CBCT) از ناحیه مفصل گیجگاهی (TMJ) بیماران با مشکلات مرتبط با اختلالات مفصل گیجگاهی-فکی (TMD) است. این تصاویر از مرکز رادیولوژی فک و صورت جمع‌آوری می‌شوند.

تصاویر CBCT از بیماران مبتلا به TMD که به مراکز رادیولوژی مراجعه کرده‌اند، جمع‌آوری خواهد شد. تعداد تصاویر حداقل ۵۰۰ تصویر از بیماران مختلف خواهد بود تا تنوع داده‌ها و قابلیت تعمیم مدل تضمین شود. تصاویر جمع‌آوری شده شامل تصاویر با کیفیت‌های مختلف خواهند بود تا مدل بتواند با انواع داده‌ها سازگار شود.

پس از جمع‌آوری، بررسی اولیه تصاویر توسط رادیولوژیست‌ها برای اطمینان از کیفیت و عدم وجود اشکالات فنی انجام خواهد شد. تصاویر با وضوح ناکافی، نویز بالا یا تصاویر ناقص حذف خواهند شد.

در مرحله پیش‌پردازش، تصاویر نرمال‌سازی، برش و تغییر اندازه خواهند شد تا به اندازه x224224 پیکسل برسند که با ورودی مدل VGG-16 سازگار باشد. علاوه بر این، تکنیک‌های افزایش داده (Data Augmentation) مانند چرخش، تغییر مقیاس و تغییرات روشنایی برای افزایش تنوع داده‌ها و جلوگیری از overfitting اعمال خواهند شد.

هر تصویر CBCT توسط یک رادیولوژیست فک و صورت بررسی شده و گزارشی متنی درباره وضعیت مفصل گیجگاهی تهیه خواهد شد. این گزارش‌ها شامل توصیفات دقیقی از ناهنجاری‌ها، ساختارهای استخوانی و دیگر ویژگی‌های مرتبط با TMD خواهند بود. گزارش‌های متنی به صورت فایل‌های متنی در کنار هر تصویر ذخیره می‌شوند.

۸. زمان‌بندی(Timeline)

ماه اول: جمع‌آوری داده‌ها

- جمع‌آوری تصاویر CBCT از مفصل گیجگاهی (TMJ) از بیماران مختلف
- بررسی و تایید کیفیت تصاویر توسط رادیولوژیست‌ها
- ایجاد فایل‌های متنی شامل گزارش‌های رادیولوژی برای هر تصویر.
- حذف تصاویر با کیفیت پایین.
- آماده‌سازی داده‌ها برای مدل یادگیری عمیق.
- اعمال تکنیک‌های افزایش داده (Data Augmentation) برای تنوع بخشیدن به مجموعه داده‌ها.

ماه دوم: طراحی و آموزش مدل

- طراحی و پیاده‌سازی مدل CNN (VGG-16) برای استخراج ویژگی‌ها از تصاویرCBCT
- آموزش مدل VGG-16 با استفاده از داده‌های آموزشی.
- طراحی و پیاده‌سازی مدل LSTM برای تولید گزارش‌های متنی.
- آموزش مدل LSTM با استفاده از ویژگی‌های استخراج‌شده از تصاویر و گزارش‌های متنی مرتبط
- ارزیابی عملکرد مدل با استفاده از داده‌های آزمون.
- تنظیم پارامترهای مدل برای بهبود دقت و کارایی.
- اعمال تکنیک‌های کاهش over-fitting مانند Dropout و.Batch Normalization

ماه سوم: تجزیه و تحلیل نتایج

- تحلیل نتایج ارزیابی مدل و بررسی معیارهای ارزیابی مانندBLEU ، ROUGE، METEORو CIDEr.
- مقایسه نتایج با پژوهش‌های قبلی و بررسی نقاط قوت و ضعف مدل پیشنهادی.

ماه چهارم: نوشتن پایان‌نامه

- نوشتن بخش‌های مختلف پایان‌نامه شامل مقدمه، مرور ادبیات، روش تحقیق، نتایج و بحث.
- بازبینی و ویرایش پایان‌نامه.
- ارسال پایان‌نامه به اساتید راهنما و مشاور برای بررسی و دریافت بازخورد.
- نگارش مقاله و ارسال آن به مجله

ماه پنجم: اصلاحات نهایی و دفاع

- اعمال تغییرات و اصلاحات نهایی بر اساس بازخورد دریافتی.
- آماده‌سازی برای جلسه دفاع.
- برگزاری جلسه دفاع و ارائه پایان‌نامه.

منابع(References)

[۱] H. Zhang and Y. Qie, "Applying Deep Learning to Medical Imaging: A Review," *Applied Sciences,* vol. 13, no. 18, p. 10521, 2023. [Online]. Available: https://www.mdpi.com/2076-3417/13/18/10521.

[۲] E. Venkatesh and S. V. Elluru, "Cone beam computed tomography: basics and applications in dentistry," *Journal of istanbul University faculty of Dentistry,* vol. 51, no. 3 Suppl 1, pp. 102-121, 2017.

[۳] M. M. A. Monshi, J. Poon, and V. Chung, "Deep learning in generating radiology reports: A survey," *Artificial Intelligence in Medicine,* vol. 106, p. 101878, 2020.

[۴] J. Valente, J. António, C. Mora, and S. Jardim, "Developments in image processing using deep learning and reinforcement learning," *Journal of Imaging,* vol. 9, no. 10, p. 207, 2023.

[۵] O. A. Montesinos López, A. Montesinos López, and J. Crossa, "Fundamentals of artificial neural networks and deep learning," in *Multivariate statistical machine learning methods for genomic prediction*: Springer, 2022, pp. 379-425.

[۶] D. Garikapati and S. Sudhir Shetiya, "Autonomous Vehicles: Evolution of Artificial Intelligence and Learning Algorithms," *arXiv e-prints,* p. arXiv: 2402.17690, 2024.

[۷] R. M. Summers, "Progress in fully automated abdominal CT interpretation," *American Journal of Roentgenology,* vol. 207 ,no. 1, pp. 67-79, 2016.

[۸] X. Wang *et al.*, "Unsupervised category discovery via looped deep pseudo-task optimization using a large scale radiology image database," *arXiv preprint arXiv:1603.07965,* 2016.

[۹] Y. Dong, Y. Pan, J. Zhang, and W. Xu, "Learning to read chest X-ray images from 16000+ examples using CNN," in *2017 IEEE/ACM international conference on connected health: applications, systems and engineering technologies (CHASE)*, 2017: IEEE, pp. 51-57 .

[۱۰] P. Rajpurkar *et al.*, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," *arXiv preprint arXiv:1711.05225,* 2017.

[۱۱] H. Wang and Y. Xia, "Chestnet: A deep neural network for classification of thoracic diseases on chest radiography," *arXiv preprint arXiv:1807.03*.۲۰۱۸ ,۰۵۸

[۱۲] J. Rubin, D. Sanghavi, C. Zhao, K. Lee, A. Qadir, and M. Xu-Wilson, "Large scale automated reading of frontal and lateral chest x-rays using dual convolutional neural networks," *arXiv preprint arXiv:1804.07839,* 2018.

[۱۳] B. Jing, P. Xie ,and E. Xing, "On the automatic generation of medical imaging reports," *arXiv preprint arXiv:1711.08195,* 2017.

[۱۴] Y. Xue *et al.*, "Multimodal recurrent model with attention for automated radiology report generation," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I*, 2018: Springer, pp. 457-466 .

[۱۵] X. Wang, Y. Peng, L. Lu, Z. Lu, and R. M. Summers, "Tienet: Text-image embedding network for common thorax disease classification and reporting in chest x-rays," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 9049-9058 .

[۱۶] Y.-C. Chang *et al.*, "Deep multi-objective learning from low-dose CT for automatic lung-RADS report generation," *Journal of Personalized Medicine,* vol. 12, no. 3, p. 417, 2022.

[۱۷] D. Sharma, C. Dhiman, and D. Kumar, "FDT− Dr2T: a unified Dense Radiology Report Generation Transformer framework for X-ray images," *Machine Vision and Applications,* vol. 35, no. 4, pp. 1-13, 2024.

[۱۸] Y. Liao, H. Liu, and I. Spasić, "Deep learning approaches to automatic radiology report generation: A systematic review," *Informatics in Medicine Unlocked,* vol. 39, p. 101273, 2023.

[۱۹] T. Syeda-Mahmood *et al.*, "Chest x-ray report generation through fine-grained label learning," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part II 23*, 2020: Springer, pp. 561-571 .

[۲۰] K.-U. Lewandrowski *et al.*, "Reliability analysis of deep learning algorithms for reporting of routine lumbar MRI scans," *International Journal of Spine Surgery,* vol. 14, no. s3, pp. S98-S107, 2020.

[۲۱] K. Yan, Y. Peng, Z. Lu, and R. M. Summers, "Fine-grained lesion annotation in CT images with knowledge mined from radiology reports," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019: IEEE, pp. 285-288 .

[۲۲] S. Loveymi, M. H. Dezfoulian, and M. Mansoorizadeh, "Automatic generation of structured radiology reports for volumetric computed tomography images using question-Specific deep feature extraction and learning," *Journal of Medical Signals & Sensors,* vol. 11, no. 3, pp. 194-207, 2021.

[۲۳] H.-C .Shin, K. Roberts, L. Lu, D. Demner-Fushman, J. Yao, and R. M. Summers, "Learning to read chest x-rays: Recurrent neural cascade model for automated image annotation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016 ,pp. 2497-2506 .

[۲۴] A. Gasimova, "Automated enriched medical concept generation for chest X-ray images," in *Interpretability of Machine Intelligence in Medical Image Computing and Multimodal Learning for Clinical Decision Support: Second International Workshop, iMIMIC 2019, and 9th International Workshop, ML-CDS 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Proceedings 9*, 2019: Springer, pp. 83-92 .

[۲۵] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[۲۶] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation,* vol. 9, no. 8, pp. 1735-1780, 1997.
[۲۷] G. Van Houdt, C. Mosquera, and G. Nápoles, "A review on the long short-term memory model," *Artificial Intelligence Review,* vol. 53, no. 8, pp. 5929-5955, 2020.