

ترکیب شبکه عصبی موبایل نت با الگوهای دودویی محلی برای تولید گزارش رادیولوژی تصاویر سی تی اسکن کبد

سمیرا لویمی^۱، میرحسین دزفولیان^۲ و محرم منصوری زاده^۲

چکیده

در دنیای امروز پزشکی توسعه روزافزون ابزار تولید تصاویر رادیولوژی پزشکی در مراکز درمانی، ایجاد سیستم‌های سبک، قابل حمل و دقیق جهت تحلیل و آنالیز تصاویر و استخراج اطلاعات تخصصی از این تصاویر را ضروری ساخته است. در بسیاری موارد تصاویر پزشکی فاقد برجسب یا حاشیه‌نویسی با اطلاعات تخصصی و کلینیکال هستند. از این رو طراحی سیستم‌هایی برای تولید اطلاعات تخصصی در مورد محتوای تصاویر یکی از چالش‌های مطرح است. در این پژوهش سیستم تولید گزارش رادیولوژی ساخت یافته مبتنی بر روش‌های حاشیه‌نویسی ارائه شده است. از جمله چالش‌های اساسی در این زمینه استخراج ویژگی‌ها و توصیف‌گرهای مناسب از تصاویر به منظور مدل‌سازی مفاهیم و محتوای تصاویر است. بدین منظور با توجه به کارآمدی فرایند یادگیری عمیق و قابلیت آن در استخراج ویژگی متناسب با هدف، در این مقاله از شبکه‌های عمیق موبایل نت به دلیل سبک و دقیق بودن، استفاده شده است. همچنین با توجه به کم بودن داده‌های آموزشی در حوزه‌های تخصصی پزشکی علاوه بر بهره‌گیری از روش‌های کاهش بیش‌برازش در شبکه موبایل نت، روش ترکیبی مبتنی بر توصیف‌گرهای عمیق و الگوی دودویی محلی ارائه شده است. نتایج بیانگر موثر بودن روش پیشنهادی هیبریدی در بهبود دقت سیستم بوده و دقت نهایی سیستم ۹۱٫۴٪ است.

کلیدواژه‌ها

یادگیری عمیق، موبایل نت، الگوی دودویی محلی، حاشیه‌نویسی تصاویر، گزارش رادیولوژی، تصاویر سی تی

۱ مقدمه

تصاویر استخراج کنند. در این راستا می‌توان سیستم‌های حاشیه‌نویسی [۱]، دسته‌بندی [۲]، بازیابی مبتنی بر محتوا، بازیابی مبتنی بر معنا [۳] و [۴] را در حوزه‌ی تصاویر پزشکی نام برد. تولید و طراحی تمامی این سیستم‌ها با بهره‌گیری از روش‌های یادگیری ماشین، و پردازش تصویر انجام می‌شود.

در [۵]، به تحلیل و بررسی سیستم‌های مختلف ارائه شده در زمینه‌ی تشخیص^۱ و یا پیش‌بینی^۲ مسائل پزشکی در مورد بیماران پرداخته شده است. از جمله این سیستم‌های می‌توان به دسته‌بندی یا حاشیه‌نویسی تصاویر پزشکی بر اساس مدالیته‌های مختلف [۱] و [۶]، تشخیص ارگان موجود در تصویر [۲] و یا تشخیص بیماری [۷] و [۸] اشاره کرد. بررسی‌ها در [۵]، نشان می‌دهد در میان روش‌های مختلف ارائه شده نحوه‌ی کاوش تصویر و استخراج ویژگی‌های بصری نقش کلیدی را دارد، که از این منظر سیستم‌های ارائه شده به سه دسته تقسیم شده است:

امروزه در حوزه‌ی سلامت و درمان، تصویربرداری پزشکی از طریق هموار کردن راه برای تشخیص سریع و قابل اعتماد، برنامه‌ریزی درمان و تجزیه و تحلیل واکنش به درمان نقش مهمی ایفا می‌کند. با ظهور سیستم‌های تصویربرداری پزشکی و به‌کارگیری بیشتر آن‌ها، تعداد تصاویر پزشکی به‌طور مداوم در حال افزایش است. این امر نیاز به ساختن سیستم‌های قدرتمندی را ایجاد می‌کند که بتوانند به‌طور خودکار و موثر چنین داده‌هایی را پردازش کرده و اطلاعات مفید و مرتبط برای پزشکان و بیماران را از

این مقاله در شهریورماه ۱۳۹۸ دریافت، در فروردین‌ماه ۱۳۹۹ بازنگری و در خردادماه پذیرفته شد.

^۱ دانشجوی دکتری مهندسی کامپیوتر، دانشگاه بوعلی سینا، همدان

رایانامه: s.loveymi@basu.ac.ir

^۲ گروه مهندسی کامپیوتر، دانشگاه بوعلی سینا، همدان

رایانامه: dezfoulia@basu.ac.ir

mansoorm@basu.ac.ir

¹ Diagnostic

² Prognostic

ساخت یافته^۲ از تصاویر پزشکی مطرح می‌شود. در واقع در تولید این سیستم هدف کشف خودکار اطلاعاتی است که یک رادیولوژیست با دیدن یک تصویر تشخیص می‌دهد و در گزارش تصویر بیمار قید می‌کند.

یکی از رویکردهایی که تا به حال در تولید گزارش رادیولوژی ساخت یافته در مورد تصاویر سی تی کمتر به آن پرداخته شده است استفاده از روش‌های یادگیری عمیق است. از طرفی در این حوزه که به کشف مسائل تخصصی در مورد تصاویر پزشکی می‌پردازد، تصاویر آموزشی در حجم بالا موجود نیست. بنابراین در راستای استفاده از روش یادگیری عمیق و در عین حال غلبه بر کم بودن تصاویر آموزشی تخصصی در سیستم پیشنهادی علاوه بر انتخاب موبایل نت به عنوان شبکه‌ای با پارامتر کمتر نسبت به شبکه‌های پیچشی دیگر و بهره بردن از روش‌های پیشگیری از بیش‌برازش، از ویژگی دستی الگوی دودویی محلی و ترکیب آن با ویژگی‌های عمیق استفاده کرده‌ایم. همچنین برای طبقه‌بندی تصاویر از روش‌های کلاسیک یادگیری ماشین مبتنی بر ترکیب طبقه‌بندها بهره برده‌ایم.

ساختار ادامه مقاله به این صورت است: در بخش بعدی کارهای مرتبط پیشین مرور و بررسی شده‌اند. سپس روش ارائه شده در این مقاله به منظور تولید خودکار گزارش رادیولوژی ساخت یافته در مورد تصاویر سی تی به تفصیل در بخش سوم ارائه شده است، همچنین مجموعه آموزشی مورد استفاده در این بخش معرفی می‌شود. سپس در بخش چهارم به تشریح مراحل آموزش شبکه، چگونگی تنظیم پارامترها و همچنین ارزیابی و تحلیل نتایج پرداخته می‌شود. در نهایت در بخش پایانی مطالب ارائه شده جمع‌بندی خواهند شد.

۲ مروری بر کارهای پیشین

در راستای مطالعه سیستم‌های ارائه شده در تولید خودکار گزارش رادیولوژی ساخت یافته و حاشیه‌نویسی تصاویر، در ادامه سیستم‌های مبتنی بر ویژگی‌های دستی، یادگیری عمیق و همچنین روش‌های ترکیبی معرفی خواهند شد.

سیستم ارائه شده در [۹]، برای حاشیه‌نویسی تصاویر کبد از تکنیک‌های بازیابی مبتنی بر محتوا (CBIR) استفاده شده تا بتواند تصاویری که مشابه با تصویر ورودی هستند را یافته و بر اساس برجسب‌های آنها تصویر ورودی را حاشیه‌نویسی کند. ویژگی‌های شکل، بافت، شدت پیکسل‌ها و BoF مبتنی بر SIFT، را استخراج کرده و از دو مرحله SVM و روش نزدیک‌ترین همسایه وزن دهی شده برای هر بخش از حاشیه‌نویسی تصاویر استفاده شده است. در این روش مدل نزدیک‌ترین همسایه بر اساس ویژگی‌های استخراج شده و همچنین با استفاده از الگوریتم انتخاب ویژگی متوالی روبه‌جلو، تولید شده است. بهترین نتایج با

۱. مبتنی بر ویژگی‌های دستی: در دسته اول دو مرحله‌ی کلیدی وجود دارد؛ ابتدا انتخاب ویژگی‌های از پیش طراحی و مهندسی شده یا کشف ویژگی‌های جدید به منظور استخراج از تصویر و سپس انتخاب روش‌های مناسب برای آموزش مدل. در این دسته شناخت دقیق از دامنه‌ی تصاویر و مواردی که باید پیش‌بینی شود بسیار مهم است.

۲. مبتنی بر یادگیری عمیق: به منظور طراحی سیستم‌های پزشکی مبتنی بر یادگیری عمیق که امروزه مورد استقبال قرار گرفته‌اند، سه منظر باید در نظر گرفته شوند: (۱) ورودی شبکه، (۲) استفاده از شبکه‌های از پیش آموزش داده شده یا آموزش از پایه، (۳) نوع شبکه. انتخاب هر کدام از این موارد به نوع مسئله‌ی پزشکی، نوع تصاویر و حجم اطلاعات در دسترس از بیماران بستگی دارد.

۳. راه‌حل‌های ترکیبی: گروهی از سیستم‌ها با هدف ترکیب ویژگی‌های دستی و عمیق تولید شده‌اند تا از مزایای هر دو روش استفاده کنند و برخی معایب را برطرف کنند.

روش‌های مبتنی بر ویژگی‌های دستی، که از میان ویژگی‌های مختلف تصویر تعدادی را استخراج و بر اساس آنها مدل‌ها را آموزش می‌دهند ممکن است در صورت تغییر ماهیت تصاویر نیاز به تغییر مجموعه ویژگی‌های استخراج شده یا روش‌های یادگیری وجود داشته باشد تا بتوان مجدداً سیستمی با دقت قابل قبول تولید کرد. در استفاده از شبکه‌های عمیق و حالت انتها-به-انتهای آنها که قدرت تعمیم بهتری دارند می‌توان تنها با تغییر مجموعه آموزش، مدل را برای هدف دیگری مانند تغییر ارگان مورد بررسی، آموزش داد و تولید کرد. در واقع یادگیری عمیق مشکل انتخاب و پیدا کردن بهینه‌ترین ویژگی‌ها از میان ویژگی‌های مختلف قابل استخراج از تصویر را حل کرده است. پیش‌نیاز اصلی این روش‌ها داده‌ی آموزشی بزرگ است که در بسیاری از موارد پزشکی، به‌ویژه در حوزه‌های تخصصی، این حجم از اطلاعات برجسب دار در اختیار نیست. از طرف دیگر به دلیل یادگیری خودکار ویژگی‌ها، شبکه‌های عمیق به‌سختی می‌توانند از هیوریستیک (اطلاعات دامنه) برای هدایت یا تغییر نحوه‌ی استخراج ویژگی برای هر کار خاص استفاده کنند. روش‌های ترکیبی مبتنی بر ترکیب ویژگی‌های عمیق و دستی به منظور استفاده از مزایای هر دو دسته و رفع معایب هر کدام ارائه شده‌اند.

به منظور انتخاب هر کدام از روش‌های بالا ابتدا می‌بایست حوزه‌ی کاری تصاویر پزشکی، چالش‌ها و اهداف مدنظر در تولید یک سیستم تحلیل تصاویر پزشکی مورد بررسی قرار گیرد. با پیشرفت کیفیت تولید سیستم‌های حاشیه‌نویسی و دسته‌بندی تصاویر یکی از چالش‌های موجود، که در این مقاله به ارائه راه‌حل برای آن می‌پردازیم، حاشیه‌نویسی تصاویر پزشکی با اطلاعات تخصصی و کلینیکال است که در قالب تولید خودکار گزارش رادیولوژی

² Structured radiology report

¹ End-to-end

تشخیص پاتولوژی تصاویر رادیوگرافی قفسه سینه که به صورت مسئله‌ی چندبرچسبی^{۱۰} است، از شبکه‌ی رزنت^{۱۱} استفاده کرده است. شبکه‌ی رزنت با سه نوع پیکربندی آموزش داده شده است، دو حالت از روش انتقال یادگیری^{۱۲} با و بدون استفاده از تنظیم دقیق^{۱۳} روی وزن‌های شبکه پیش آموزش^{۱۴} دیده، و آموزش از پایه^{۱۵}. البته آخرین لایه متراکم معماری اصلی با یک لایه متراکم جدید با تطبیق تعداد برچسب‌ها و اضافه کردن تابع فعال‌سازی سیگموئید^{۱۶} برای مسئله چند برچسب، جایگزین شده است. بهترین نتایج با استفاده از رزنت ۳۸ که از پایه آموزش داده شده به دست آمده که این موضوع با توجه به اینکه دیتاست مورد استفاده بالای صد هزار نمونه دارد، دور از انتظار نیست. نویسندگان در [۱۵]، با هدف دستیابی به سیستم بی‌درنگ^{۱۷} و تشخیص دقیق نقص سطوح از روش‌های یادگیری عمیق مبتنی بر موبایل نت استفاده کرده‌اند. ویژگی‌های لایه‌های مختلف پیچشی^{۱۸} موبایل نت به شبکه عمیق^{۱۹} SSD برای رگرسیون و رگرسیون جعبه محدود^{۲۰} داده شده است. مدل MobileNet - SSD می‌تواند به شدت تعداد پارامترها را کاهش داده، و در شرایط سخت‌افزاری محدود به دقت بالاتری دست یابد. مدل کامل شامل چهار بخش است: لایه ورودی برای وارد کردن تصویر هدف، شبکه مبنای پیش‌آموزش داده و تنظیم دقیق شده‌ی موبایل نت برای استخراج ویژگی‌های تصویر، SSD برای رگرسیون طبقه‌بندی و رگرسیون جعبه محدود و لایه خروجی برای ارائه نتایج تشخیص.

در [۶]، طبقه‌بندی ماهیت^{۲۱} و تشخیص مفاهیم^{۲۲} تصاویر پزشکی با استفاده از روش‌های یادگیری عمیق و با تکیه بر انتقال یادگیری بررسی شده است. برای طبقه‌بندی ماهیت، عملکرد ۷ شبکه‌ی VGG_16 [۱۶]، VGG-19 [۱۶]، ResNet-50 [۱۷]، Inception-v3 [۱۸]، Xception [۱۹]، MobileNet [۲۰] و Inception-ResNet-v2 [۲۱]، مقایسه شده است. به این صورت که هر شبکه روی ImageNet [۲۲]، پیش‌آموزش داده شده و لایه‌ی تماماً متصل آن روی تصاویر پزشکی آموزش داده شده است، سپس از این شبکه به منظور استخراج ویژگی استفاده می‌شود و یادگیری با استفاده از یک دسته‌بندی کننده رگرسیون لجستیک انجام می‌شود. نتایج نشان می‌دهد بهترین نتایج را Inception-v3 و سپس موبایل نت داشته‌اند. در بخش تشخیص مفاهیم این مقاله روش‌های مبتنی بر استخراج ویژگی-

استفاده از ویژگی‌های BoF و نزدیک‌ترین همسایه به دست آمده است. [۱۰]، دو روش کلی برای حاشیه‌نویسی تصاویر کبد استفاده شده است، در روش اول ویژگی‌های شکل و بافت با استفاده از روش هارالیک و موجک گابور استخراج و آموزش مدل از روش دسته‌بندی کننده‌ی جنگل تصادفی^۱ و برای برخی سوالات روش نزدیک‌ترین همسایه انجام می‌شود. روش دوم از امضای ویژه^۲ کبد استفاده می‌کند. در این روش یک برش از سی‌تی اسکن سه‌بعدی کبد انتخاب و سپس در یک بلوک مستطیلی با ابعاد ثابت نرمال شده و به بلوک‌های کوچک تقسیم شده است. بعد از اعمال فیلتر یک‌بعدی Log-Gabor، فاز غالب هر بلوک استخراج و به چهار سطح تقسیم می‌شود تا الگوی منحصر به فرد کبد را کدگذاری کند. فاصله همینگ برای بازیابی به کار گرفته شده است. نویسندگان مقاله بهترین نتایج را در ImageCLEF 2015 کسب کرده‌اند. در [۱۱]، از ترکیب ویژگی‌های بافت شامل BMED و تبدیل موجک گابور برای حاشیه‌نویسی تصاویر کبد استفاده شده است. به این صورت که با استفاده از روش تجزیه عملکرد حالت ذاتی IMF^۳ تصویر را تجزیه کرده و از هر کدام ویژگی‌های بافت تبدیل موجک گابور را استخراج می‌کند و برچسب‌های تصویر ورودی بر اساس ۵ نزدیک‌ترین همسایه انتخاب می‌شود. در [۱۲]، با تمرکز بر ضایعات کبدی سیستمی برای بازیابی تصاویر مبتنی بر حاشیه‌نویسی تصاویر ارائه شده است. در این مقاله برای هر ضایعه ویژگی‌های هیستوگرام تصویر مبتنی بر مقادیر خاکستری (هانسفیلد^۴) و بردار وزنی ارتباط بین برچسب‌های مختلف و ویژگی‌ها با استفاده از روش‌های دسته‌بندی دودویی SVM^۵ و LR^۶ و LD^۷ استخراج می‌شود. نویسندگان در [۱۳]، از ویژگی‌های بافت، شکل و گابور و یک روش پیشنهادی استخراج ویژگی با نام DLBP، در تولید گزارش رادیولوژی ساخت یافته تصاویر کبد، روی دیتاست ImageCLEF 2015، استفاده کرده‌اند و فرآیند آموزش مدل‌ها با الگوریتم طبقه‌بند ترکیبی انجام شده است. [۲]، به منظور تشخیص ارگان موجود در هر اسلایس از تصویر سی‌تی سه‌بعدی بدن با تقسیم‌بندی کلی مغز، گردن، قفسه سینه، شکم و لگن از روش‌های یادگیری عمیق مبتنی بر شبکه‌های پیچشی استفاده شده است. برای هر ارگان یک دیتاست جداگانه تولید شده و برای هر کدام شبکه‌های الکسنت^۸ و گوگلنت^۹ استفاده شده است طبق گزارش مقاله به طور کلی زمان آموزش الکسنت کمتر از گوگلنت و دقت تشخیص برخی ارگان‌ها با استفاده از الکسنت و در برخی گوگلنت بالاتر است. [۱۴]، برای

¹⁰ Multi label problem

¹¹ ResNet

¹² Transfer learning

¹³ Fine-tuning

¹⁴ Pre-trained

¹⁵ Scratch

¹⁶ Sigmoid

¹⁷ Real time

¹⁸ Convolutional

¹⁹ Single Shot MultiBox Detector

²⁰ Boundary box regression

²¹ Modality classification

²² Concept detection

¹ Random forest

² Specific signature of the liver

³ Intrinsic Mode Function decomposition

⁴ Hounsfield

⁵ Support Vector Machine

⁶ Logistic Regression

⁷ Linear Discriminant

⁸ AlexNet

⁹ GoogLeNet

آموزش مدل‌های طبقه‌بندی کننده و روند دوم استفاده از روش‌های یادگیری عمیق که اخیراً بسیار مورد استفاده قرار می‌گیرد. سوالات موجود در یک گزارش رادیولوژی در مورد تصویر یک بیمار گستردگی موضوعی دارند و تخصصی و متنوع هستند زیرا در مورد بخش‌های مختلف از جمله عروق، بافت‌ها و ابعاد اجزای موجود در تصویر و پاتولوژی ضایعات می‌باشند. بنابراین انتخاب ویژگی‌های که باید از هر تصویر استخراج شود تا بتواند توصیفگرهای مناسب برای آموزش مدل‌های رده‌بندی و یافتن پاسخ هر سوال باشد، یکی از چالش‌های موجود است. همان‌طور که یک رادیولوژیست با توجه به ماهیت هر سوال به ویژگی‌های مختلف از لحاظ محتوایی و ظاهری توجه می‌کند. یکی از مزایای استفاده از روش‌های یادگیری عمیق در این مسئله این است که می‌توان برای هر سوال یک شبکه end-to-end آموزش داد که بخش front-end که می‌توان آن را به‌عنوان بخش استخراج ویژگی در نظر گرفت برخلاف روش‌های کلاسیک به‌صورت کاملاً هم‌زمان با بخش back-end که وظیفه‌ی طبقه‌بندی را به عهده دارد، بهینه‌سازی می‌شوند.

از طرفی به‌منظور استفاده از روش‌های یادگیری عمیق در حوزه‌ی پزشکی باید دو چالش را مدنظر قرار داد و آنها را تا جای ممکن حل کرد: ۱) در دنیای پزشکی نیاز به سیستم‌هایی داریم که سبک باشند و بتوانند در محل‌هایی مانند مطب‌ها و کلینیک‌ها که ممکن است از لحاظ سخت‌افزاری و محاسباتی امکانات کمتری^۵ داشته باشند و یا حتی تلفن‌های همراه بیماران یا پزشکان نصب شوند و اجرا و آنالیز را انجام دهند، ۲) تعداد داده‌های آموزشی در حوزه‌ی های پزشکی و مخصوصاً در حوزه‌های تخصصی مانند کاری که قرار است در این مقاله انجام شود کم است و شبکه‌های عمیق نیاز به تعداد داده‌ی آموزشی بالا دارند. برای حل مورد اول شبکه‌های موبایل نت را انتخاب کردیم که در ادامه به معرفی آنها می‌پردازیم و برای مسئله دوم از روش‌های پیش‌پردازش و تقویت داده بهره بردیم. همچنین در راستای تکمیل نمایشی^۶ که شبکه عمیق از تصاویر به دست می‌آورد، ویژگی دستی را نیز مورد استفاده قرار داده‌ایم و هر دو حالت استفاده از شبکه عمیق به‌صورت end-to-end (شکل ۱) و ترکیب ویژگی‌های استخراج شده از طریق شبکه‌ی عمیق با ویژگی‌های دستی (شکل ۲) و سپس طبقه‌بندی به کمک روش‌های مبتنی بر ترکیب طبقه‌بندی را در حوزه‌های دوعده‌ی و سه‌بعدی مورد بررسی قرار داده‌ایم که در ادامه به تفصیل اجزای سیستم معرفی می‌شوند.

۳-۱ مجموعه آموزشی و پیش‌پردازش‌های انجام شده

مجموعه آموزشی استفاده‌شده در این مقاله Liver ct Annotation ImageCLEF 2015 [۲۶] می‌باشد که با هدف تولید یک گزارش رادیولوژی ساخت‌یافته در مورد تصویر سی تی

های کلاسیک شامل LBP الگوی دودویی محلی و GLCM ماتریس همجواری، با ویژگی‌های استخراج‌شده از شبکه عمیق Inception-v3 مقایسه شده است که نتایج بیانگر بهتر بودن روش‌های مبتنی بر انتقال یادگیری عمیق می‌باشد. نتایج در هر دو مقوله نشان می‌دهد که روش‌های انتقال یادگیری عمیق در حوزه پزشکی که در آن داده‌های آموزشی کمیاب است موثر می‌باشد. مقاله [۲۳]، به‌منظور دسته‌بندی تصاویر بیومدیکال از ترکیب ویژگی‌های دستی و عمیق استفاده‌شده است. ویژگی‌های عمیق از سه شبکه‌ی عمیق پیچشی Caffe-ref [۲۴]، VGG-f و VGG-19 که روی ImageNet پیش‌آموزش و روی دیتاست آموزشی تنظیم شده‌اند، و ویژگی‌های دستی با استفاده از LBP و BoF، استخراج می‌شود. از PCA برای کاهش ابعاد و از دسته‌بندی کننده AdaBoost با شبکه عصبی BP استفاده‌شده است. نتایج بیانگر بهبود دقت اندازه‌گیری در حالت ترکیب ویژگی‌های دستی و عمیق مبتنی بر انتقال یادگیری، می‌باشد. لازم به ذکر است دیتاست مورد استفاده ImageCLEF 2016 برای دسته‌بندی تصاویر می‌باشد که نسبتاً کوچک است. در [۲۵]، ترکیب روش‌های یادگیری عمیق و کلاسیک با هدف طبقه‌بندی ماهیت تصاویر بیومدیکال روی دیتاست ImageCLEF در سال‌های ۲۰۱۳ و ۲۰۱۶ ارائه‌شده و ترکیب با تاخیر نامیده شده است. روش کار به این صورت است که در بخش عمیق، تصاویر به شبکه‌های پیچشی AlexNet، VGG16، VGG19، پیش-آموزش داده‌شده روی ImageNet داده‌اند و خروجی لایه‌ی تولیدشده است، از طرفی ویژگی‌های BoVW از تصاویر استخراج و به SVM داده‌شده است، سپس خروجی شبکه‌ها با خروجی احتمال خلفی^۱ SVM با استفاده از روش‌هایی مثل ماکزیمم ترکیب شده است.

۳ روش پیشنهادی

در این مقاله هدف طراحی سیستم تولید خودکار گزارش رادیولوژی ساخت‌یافته از تصاویر سی تی شکمی می‌باشد. هر گزارش، حاوی سوالاتی دو یا چندگزینه‌ای در مورد محتوای تصویر سی تی (بخش کبد) و پاتولوژی ضایعات^۲ موجود در آن و عروق می‌باشد. به‌منظور طراحی سیستمی که بتواند برای هر کدام از سوالات موجود در گزارش رادیولوژی بهترین پاسخ را بر اساس محتوای تصویر سی تی بیمار پیدا کند، روش‌های موجود در حوزه‌ی های پرسش و پاسخ بصری^۳ و حاشیه‌نویسی تصاویر^۴ به‌کاربرده می‌شود. در تولید چنین سیستم‌هایی دو روند کلی پیشرو است، استفاده از روش‌های مختلف استخراج ویژگی و به دنبال آن

¹ Posterior probability

² Lesion

³ Visual Question Answering (VQA)

⁴ Image annotation

⁵ Computationally limited platform

⁶ Representation

موبایل نت: شبکه موبایل نت با هدف بهبود عملکرد بلادرنگ شبکه‌های عصبی عمیق تحت شرایط محدود سخت‌افزاری توسعه یافته است و عملاً جز گروه شبکه‌های کارای عصبی پیچشی قرار می‌گیرد [۲۰]. این شبکه می‌تواند تعداد پارامترها را بدون پایین آمدن دقت کاهش دهد. مطالعات قبلی نشان داده نسبت پارامترهای مورد نیاز برای ساخت یک شبکه موبایل نت در مقایسه با یک شبکه Visual Geometry Group (VGG-16) برای رسیدن به دقت یکسان یک به ۳۳ است [۱۵].

ایده اصلی در موبایل نت این است که لایه‌های پیچشی، که برای انجام وظایف بینایی کامپیوتر ضروری هستند اما برای محاسبه بسیار گران هستند، می‌توانند با به اصطلاح پیچشی‌های جدا شده در عمق^۷ جایگزین شوند. که نتیجه‌ی آن کاهش سربار پردازشی است. در این روش به جای اینکه در یک عملیات پیچشی معمولی که هر فیلتر روی تمامی ماتریس‌های ورودی اعمال می‌شود و سپس نتایج همه باهم ترکیب و نهایتاً یک خروجی ارائه می‌شود، مرحله پیچشی به دو مرحله تبدیل می‌شود. یعنی ابتدا یک فیلتر روی ماتریس یا همان نگاشت ویژگی^۸ اعمال می‌شود و در مرحله‌ی دوم یک لایه پیچشی^۱ در نقطه‌ای قرار می‌گیرد که عمل ترکیب ماتریس‌ها را انجام می‌دهد. پس از هر لایه پیچشی یک الگوریتم نرمال‌سازی دسته‌ای^۹ و سپس یک تابع فعال‌سازی^{۱۰} ReLU اعمال می‌شود.

تابع فعال‌سازی مورد استفاده در موبایل نت ReLU6 است، که عملکرد آن شبیه ReLU است و از بزرگ شدن مقادیر پیشگیری می‌کند. مزیت اصلی استفاده از ReLU این است که برای همه‌ی ورودی‌های بزرگتر از صفر، یک مشتق ثابت دارد. این مشتق ثابت به یادگیری شبکه سرعت می‌بخشد. این تابع از رابطه (۱) محاسبه می‌شود:

$$y = \min(\max(x, 0), 6) \quad (1)$$

که x مقدار هر پیکسل از نگاشت ویژگی است. معماری شبکه‌ی موبایل نت به صورت مجموعه‌ای از بلاک‌ها است، که پیچشی‌های عمق و نقطه‌ای، نرمال‌سازی دسته‌ای و تابع فعال‌سازی، باهم یک بلاک را تشکیل می‌دهند (شکل ۱). معماری کامل موبایل نت شامل یک لایه پیچشی معمولی 3×3 به عنوان اولین لایه و به دنبال آن مجموعه‌ای از بلاک‌ها می‌باشد و لایه آخر یک لایه‌ی کلی حداکثر تجمع است. موبایل نت همچنین پارامتر ضریب عرض (آلفا) دارد که یک ضرب‌کننده‌ی عمق است که قابلیت تغییر تعداد کانال‌های در هر لایه را ایجاد می‌کند. تغییر شبکه و افزایش یا کاهش عرض شبکه و تعداد پارامترها در مقایسه با نسخه اصلی ارائه شده از شبکه با استفاده از تغییر ضریب آلفا امکان‌پذیر خواهد بود. در این مقاله از جدیدترین نسخه‌ی موبایل نت که نسخه‌ی شماره ۲ می‌باشد استفاده شده است.

شکمی هر بیمار مطرح شده است و حاوی سوالاتی در مورد ویژگی‌های کبد (مانند تراکم بافتی^۱، جایگذاری مناسب^۲، لوب‌های کبد^۳ و ...)، عروق (مانند شریان کبدی^۴) و ضایعات موجود در کبد (مانند شکل ظاهری و بافت ضایعات) می‌باشد. این دیتاست حاوی مجموعه‌ای از اطلاعات ۵۰ بیمار به صورت: تصویر سی‌تی شکمی سه‌بعدی بیمار، ناحیه مشخص‌کننده کبد و مستطیل مشخص‌کننده ناحیه‌ی ضایعات، شماره اسلایس‌هایی که ضایعات دیده می‌شوند و فایل rdf حاوی پاسخ سوالات گزارش رادیولوژی بیمار، می‌باشد. رویکرد بدیهی عملکرد مستقیم در حجم سه‌بعدی و آموزش شبکه‌ی سه‌بعدی می‌باشد، اما از آنجا که تعداد تصاویر موجود برای آموزش بسیار کم است و سیستم را در معرض بیش-برازش قرار می‌دهد. در این راستا هر کدام از اسلایس‌های دوبعدی به صورت جداگانه پردازش می‌شود و در واقع به عنوان یک تصویر از مجموعه آموزش مورد استفاده قرار می‌گیرد. بنابراین برای تولید دیتاست دوبعدی، اسلایس‌هایی از تصاویر سی‌تی بیماران انتخاب می‌شود که کبد و ضایعات کبد قابل مشاهده باشند و دیگر اجزای شکمی از هر اسلایس حذف می‌شوند. یکی دیگر از روش‌هایی که برای تقویت داده و پیشگیری از بیش‌برازش و حفظ کردن جزئیات دقیق تصاویر آموزشی توسط شبکه انجام شده استفاده از روش‌های افزایش داده^۵ می‌باشد. روش افزایش داده یک روش پیش‌پردازش است که باعث افزایش اندازه موثر مجموعه داده‌های آموزشی و در نتیجه آموزش بهتر و موثرتر مدل‌های یادگیری عمیق می‌شود. این روش اولین بار توسط [۲۷] به منظور بهبود و ساده‌سازی شبیه‌سازی^۶ استفاده شده است. در استفاده از این روش باید توجه داشت از روش‌هایی برای تغییر عکس استفاده کرد که عکس تولید شده خارج از فضای حالت مسئله نبوده و در شرایط واقعی مسئله وجود داشته باشد [۲۸]. در مورد تصاویر کبد استفاده از روش‌هایی مانند چرخش یا افزودن نویز به منظور افزایش داده امکان‌پذیر نیست زیرا روی کیفیت بافت و محل عروق و لوب‌های مختلف کبد که در سوالات مختلف آنالیز می‌شوند، تاثیر منفی دارد و ممکن است منجر به پاسخ‌های اشتباه و آموزش مدل‌های نادرست شود. به همین دلایل تنها روش مطمئن و انتخاب شده در این مقاله برای عدم تاثیرگذاری روی نتایج نهایی سوالات مختلف روش شیف‌ت می‌باشد.

۲-۳ استفاده از شبکه عصبی عمیق انتها-به-انتها

به منظور استفاده از این شبکه طبق شکل ۱ ابتدا تصاویر برش‌ها که دوبعدی هستند پیش‌پردازش می‌شوند و سپس به موبایل نت و لایه‌های تماماً متصل، حذف تصادفی و طبقه‌بندی وارد می‌شوند.

¹ Liver Density

² Liver Placement

³ Lobes

⁴ Hepatic Artery

⁵ Augmentation

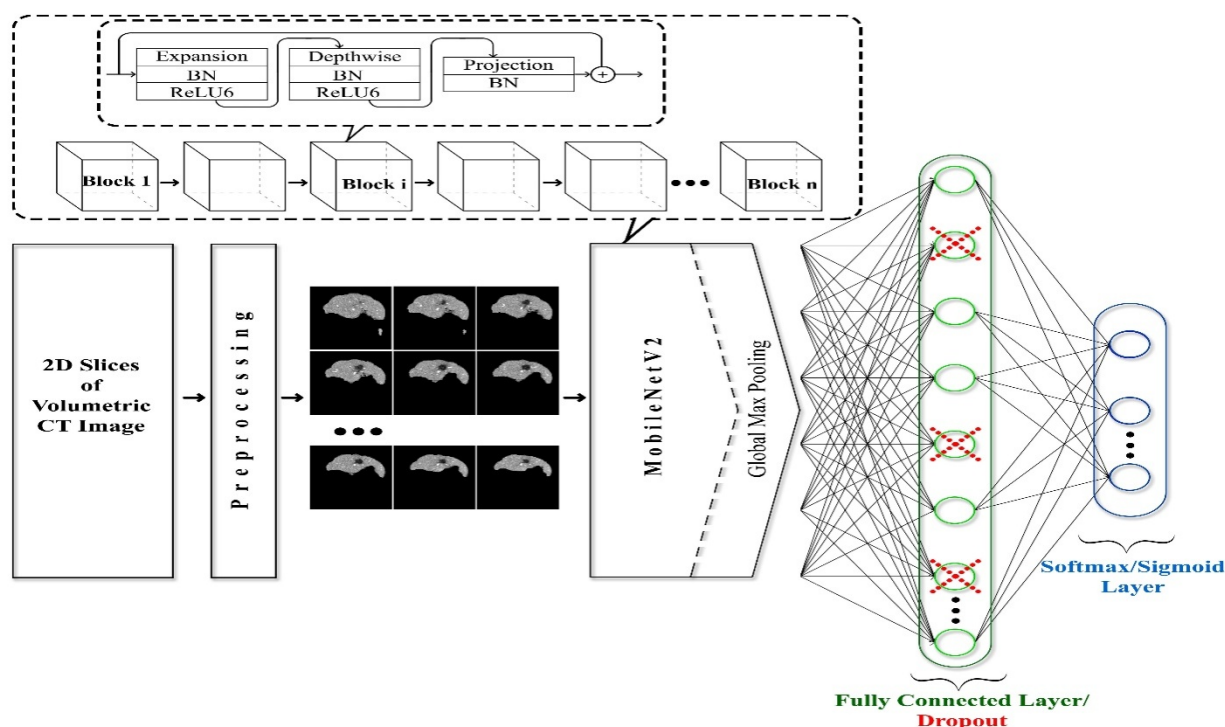
⁶ Simulation

⁷ Depthwise separable convolution

⁸ Feature map

⁹ Batch normalization

¹⁰ Rectified Linear Unit



شکل ۱ الگوی کلی سیستم طراحی شده برای هر سوال برای استفاده از شبکه عصبی موبایل نت به صورت انتها-به-انتها

دارای ۱۷ بلاک است و سپس یک پیش‌پردازش معمولی 1×1 و یک لایه کلی تجمعی حداکثر یا میانگین، که پس از آن می‌توان مانند شکل ۱ لایه‌های طبقه‌بندی را تنظیم کرد. بنابراین تفاوت اصلی در نسخه ۲ نسبت به نسخه ۱، اتصالات باقیمانده و لایه‌های گسترش و نگاشت هستند. در [۲۹]، نشان داده شده که نسخه‌ی دوم موبایل-نت پارامترهای کمتری نسبت به نسخه‌ی یک استفاده می‌کند و سرعت بیشتری در دستگاه‌های تلفن همراه دارد.

لایه تمام متصل^۶ و حذف تصادفی^۷: با توجه به اینکه هدف نهایی نهایی در این مقاله، این است که برای هر کدام از سوالات شبکه‌ای توسعه داده شود که بهترین پاسخ را بر اساس محتوای تصویر تولید کند لذا پس از موبایل نت لایه‌ی تماماً متصل تعریف می‌شود تا بتواند قابلیت طبقه‌بندی و یافتن پاسخ را ایجاد کند. نورون‌هایی که در یک لایه تماماً متصل قرار دارند با تمام نورون‌های موجود در لایه قبلی (لایه تجمعی در انتهای موبایل نت) ارتباط دارند. از آنجایی که یک لایه تماماً متصل بیشترین تعداد پارامترها را به خود اختصاص می‌دهد، در نتیجه در معرض بیش‌برازش خواهد بود. روش حذف تصادفی به منظور جلوگیری از بیش‌برازش معرفی شد [۳۱] و [۳۲]، نحوه کار آن به این صورت است که در هر مرحله از آموزش، هر نورون یا با احتمال $1-p$ (از شبکه) بیرون انداخته شده و یا با احتمال p نگه داشته می‌شود، بطوریکه نهایتاً با پرهیز از آموزش تمامی نورون‌ها بر روی تمامی داده‌ها، یک شبکه کاهش داده شده باقی بماند و همچنین باعث افزایش سرعت آموزش نیز می‌شود. [۳۳] در تحقیقات خود، بهینگی روش حذف

در نسخه ۲ موبایل نت در هر بلاک ۳ لایه پیش‌پردازش وجود دارد که هدف اصلی آنها کمتر کردن حجم محاسبات با حفظ میزان دقت است [۲۹]. لایه‌ی پیش‌پردازش عمق ورودی را فیلتر می‌کند و لایه پیش‌پردازش نقطه‌ای تعداد کانال‌ها را کوچکتر می‌کند و به آن لایه نگاشت^۱ می‌گویند زیرا داده‌های با بعد بالا (کانال‌های مختلف) را به تانسورهای با بعد کمتر تبدیل می‌کند. این دسته از لایه‌ها عموماً با نام لایه تنگنا^۲ نیز شناخته می‌شوند زیرا حجم دیتایی که از آنها عبور می‌کند را کاهش می‌دهند. جدیدترین عضو نسخه ۲ موبایل-نت لایه اول هر بلاک است که یک پیش‌پردازش ۱ در ۱ است و هدف آن افزایش تعداد کانال‌های تصویر قبل از رفتن به لایه پیش‌پردازش عمق است. در واقع این لایه عکس لایه نگاشت عمل می‌کند. میزان گسترش داده‌ها در این لایه با پارامتر عامل گسترش^۳ مشخص می‌شود که به طور پیش فرض ۶ است. دومین نوآوری در نسخه ۲ موبایل نت اتصال باقیمانده^۴ است برای کمک به حرکت گرادینت‌ها^۵ در طول شبکه، و فقط در مواردی استفاده می‌شود که تعداد کانال‌های وارد شده به بلاک برابر با تعداد کانال‌هایی که از آن خارج می‌شوند باشد، که در بعضی از بلاک‌ها رخ می‌دهد. طبق معمول هر لایه دارای نرمال‌سازی دسته‌ای و تابع فعال‌سازی ReLU6 است، البته لایه‌ی نگاشت تابع فعال‌سازی ندارد [۲۹] و [۳۰]. از آنجاکه این لایه دیتای با بعد کم تولید می‌کند توسعه‌دهندگان نسخه ۲ [۳۰] دریافتند که غیرخطی نمودن بعد از این لایه اطلاعات مفید را از بین می‌برد. نسخه‌ی ۲ موبایل نت

¹ Projection

² Bottleneck

³ Expansion factor

⁴ Residual connection

⁵ Flow of the gradients

⁶ Fully connected layer

⁷ Drop out

طبقه‌بندی و در نظر گرفتن خروجی آخرین لایه‌ی موبایل‌نت که لایه تجمعی حداکثر است، به دست آمده است. الگوی دودویی محلی به عنوان ویژگی استخراج شده به صورت دستی از هر تصویر با بردار ویژگی عمیق ترکیب شده است (شکل ۲). الگوی دودویی محلی اطلاعات بافت محلی تصویر را کدگذاری می‌کند [۳۴].

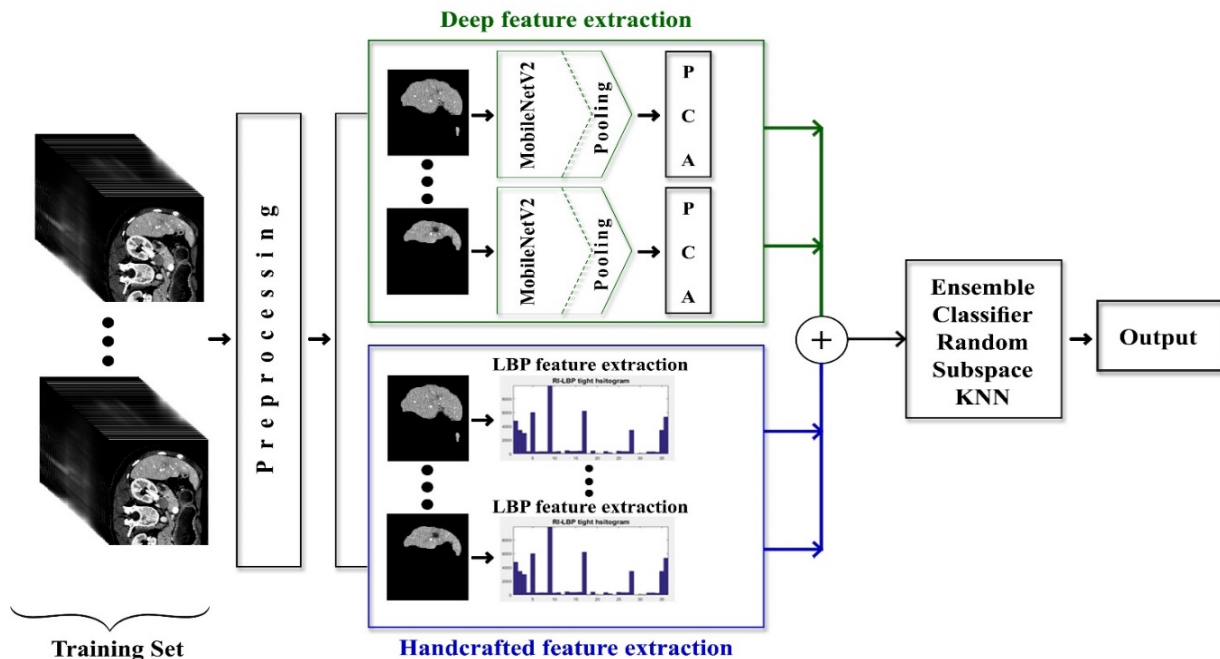
در نهایت به منظور طبقه‌بندی یا یافتن پاسخ هر سوال، ویژگی‌های ترکیبی به دسته‌بندی کننده‌ی مبتنی بر ترکیب طبقه‌بندها ارسال می‌شود تا فرآیند ساخت مدل برای هر سوال صورت گیرد. لازم به ذکر است روش پیشنهادی در شکل ۲ برای تصاویر دوبعدی می‌باشد و هر برش از تصویر سه‌بعدی سی‌تی را جداگانه بررسی می‌کنند. برای استفاده از روش پیشنهادی در حوزه‌ی سه‌بعدی در این مقاله از ویژگی‌های ترکیبی برش‌های مختلف یک تصویر سی‌تی میانگین گرفته می‌شود و به طبقه‌بند ارسال می‌گردد. بنابراین روش پیشنهادی در هر دو حوزه‌ی تصاویر دوبعدی و سه‌بعدی قابل استفاده می‌باشد.

تصادفی را مورد تحلیل قرارداد و نهایتاً عنوان کرد که یک شیوه یادگیری گروهی بسیار موثر است.

لایه طبقه‌بندی: از آنجا که لایه‌ی تماماً متصل نتیجه‌ی شبکه را در قالب یک بردار تولید کرده است می‌توان از اطلاعات این لایه برای طبقه‌بندی استفاده کرد. بنابراین در لایه‌ی طبقه‌بندی از توابع سافت‌ماکس یا سیگموئید بر اساس اینکه پاسخ سوالی که در حالت توسعه شبکه برای آن هستیم چندگزینه‌ای یا دوگزینه‌ای باشد، استفاده می‌شود.

۳-۳ ترکیب ویژگی‌های عمیق و الگوی دودویی محلی

پس از آموزش و ساخت شبکه برای هر سوال، می‌توان از لایه‌های مختلف شبکه برای استخراج ویژگی از تصویر استفاده کرد. به ویژگی‌های استخراج شده از لایه‌های مختلف شبکه ویژگی‌های عمیق گفته می‌شود. در روش پیشنهادی ویژگی‌های عمیق و دستی ترکیب می‌شوند و برای آموزش طبقه‌بندی کننده‌ی ترکیبی استفاده می‌شوند. ویژگی‌های عمیق، با حذف لایه‌های تماماً متصل و



شکل ۲: الگوی کلی سیستم هیبریدی طراحی شده برای هر سوال (ترکیب ویژگی عمیق و دستی)

آزمایش می‌شود و دقت نهایی سیستم بر اساس میانگین دقت تمامی سوالات محاسبه می‌شود.

برای آموزش موبایل‌نت با توجه به کم بودن تعداد داده‌های آموزشی در مقایسه با مجموعه‌های مورد نیاز برای آموزش شبکه‌های عمیق، روش انتقال یادگیری بر اساس ImageNet انتخاب شده است. با توجه به اینکه ImageNet برای تصاویر رنگی طراحی شده است. بنابراین ما باید مقادیر هر پیکسل از هر اسلایس از تصویر سی‌تی را سه بار کپی کنیم تا یک تصویر با ۳ کانال قرمز، سبز و آبی را شبیه‌سازی کنیم. تابع زیان استفاده شده در سوالات با پاسخ

۴ نتایج

۴-۱ تنظیم پارامترهای سیستم

به منظور بررسی دقت و کارایی سیستم پیشنهادی از معیار دقت^۱ استفاده می‌شود. برای هر سوال موجود در گزارش رادیولوژی که در مورد کبد، ضایعه‌های موجود در آن یا عروق است، مدلی جداگانه با استفاده از اعتبارسنجی متقابل^۲ آموزش داده شده و

^۱ Accuracy

^۲ Cross validation

می باشند در نظر گرفته شده است و برش هایی که کبد یا ضایعه ها در آنها مشخص نیستند حذف شده اند.

روش اول (دوبعدی_سطح-اسلایس): دقت سیستم در سه حالت مورد بررسی قرار گرفته است که شامل: (۱) استفاده از یادگیری عمیق به صورت انتها-به-انتها (شبکه عمیق موبایل نت به عنوان بخش front-end و استفاده از لایه های تماما متصل ۵۱۲ تایی و حذف تصادفی و سافت ماکس/سیگموئید در بخش back-end) (شکل ۱)، (۲) ویژگی دستی الگوی دودویی محلی و دسته بندی کننده RSKNN (شکل ۳) و روش پیشنهادی شامل ترکیب ویژگی های دستی و عمیق و دسته بندی کننده RSKNN (شکل ۲). جدول ۱ نتایج را با استفاده از روش های معرفی شده و اعتبارسنجی متقابل پنج بخشی نشان می دهد.

نتایج جدول ۱ میانگین دقت هر روش برای تمامی سوالات است در شکل ۳ دقت هر کدام از روش ها برای هر سوال به صورت جداگانه نشان داده شده است. لازم به ذکر است که در روش مبتنی بر یادگیری عمیق دقت و زیان برای مدل ساخته شده برای هر سوال در دورهای مختلف مورد بررسی و آنالیز قرار گرفته است که شکل ۴ این نمودارها را برای تعدادی از سوالات نشان می دهد.

نتایج جدول ۱ نشان دهنده برتری روش پیشنهادی نسبت به دو روش دیگر است و همچنین شکل ۳ نیز نشان می دهد در اکثریت سوالات روش هیبرید نتایج بهتری به دست آورده است.

روش دوم (سه بعدی_سطح-حجم): در این روش، برای هر بیمار در مجموعه آموزشی برش های کلیدی بر اساس اطلاعات بیمار از تصویر سی تی استخراج می شود، که تعداد این برش های کلیدی برای بیماران مختلف متفاوت است. سپس ویژگی عمیق و الگوی دودویی محلی از هر برش استخراج می شود، زیرا بر اساس نتایج به دست آمده در روش اول، بهترین دقت برای هر برش را ترکیب ویژگی عمیق و الگوی دودویی، داشته است (جدول ۱). در مرحله بعد برای تولید بردار ویژگی نهایی یک تصویر سی تی سه بعدی در سطح حجم برای یک بیمار، میانگین بردارهای ویژگی تمامی برش های کلیدی آن محاسبه می شود. از این بردارهای ویژگی با استفاده از RSKNN برای ساخت مدل هر سوال استفاده می کنیم. البته باید دقت کرد که بردار ویژگی عمیق هر تصویر متناسب با سوال مطرح شده و شبکه ی ساخته شده برای همان سوال استخراج می شود. از آنجا که مقالاتی که روی دیتاست مشابه با هدف مشابه کار کرده اند از روش دوم برای ارزیابی استفاده کرده اند نتایج روش پیشنهادی سه بعدی این مقاله، با این روش ها در جدول ۲ مقایسه شده است. روش پیشنهادی در مقایسه با اکثر روش های دیگر، با وجود استفاده از بردار ویژگی با ابعاد نسبتا کمتر، دقت بالاتری به دست آورده است. یکی از دلایل برتری روش پیشنهادی را می توان بهره گیری از روش یادگیری عمیق برای هر سوال به صورت جداگانه که نتیجه ی آن استخراج ویژگی های

دوگزینه ای؛ آنتروپی متقابل دودویی^۱ و سوالات با پاسخ چندگزینه ای؛ آنتروپی متقابل دسته ای^۲، می باشد. ما از سبک آموزشی دسته ای^۳ با اندازه دسته ۳۲ استفاده می کنیم. از آنجا که مجموعه آموزشی ما کوچک تر از یک ۲۵۰۰ ام مجموعه داده ImageNet است، ما نرخ یادگیری را به اندازه ۰,۰۰۰۱ تنظیم کردیم و هر ۵ مقطع آن را تا یک دهم کاهش می دهیم، که هدف آن جلوگیری از بیش برآزش مدل ها است. به منظور استخراج ویژگی از شبکه موبایل نت از آخرین لایه که حاوی ۱۲۸۰ داده است استفاده شده است و با استفاده از تکنیک PCA، کاهش بعد به ۱۰۰ انجام شده است. برای آموزش و تست شبکه موبایل نت از جی پی یو تسلا K80 و ۱۲ گیگابایت رم در گوگل کولب^۴ استفاده شده است. الگوی دودویی محلی با تقسیم هر تصویر به ۴ بخش و استخراج هیستوگرام مقادیر حاصل از کدگذاری الگوی دودویی محلی هر بخش با استفاده از همسایگی (P) ۱۶ و شعاع (R) ۲ تولید شده است.

برای آموزش مدلی که ورودی آن ترکیب ویژگی های حاصل از موبایل نت و الگوی دودویی است روش ترکیبی^۵ زیرمجموعه تصادفی^۶ با یادگیرنده های ضعیف نزدیک ترین k همسایه انتخاب شده است (RSKNN) که دسته بندی را بر اساس مشابه ترین تصاویر در مجموعه آموزش انجام می دهد. این انتخاب به دو دلیل انجام شده است؛ اول اینکه با توجه به بررسی مجموعه داده ها سوالات مختلف توزیع های متفاوتی روی داده های تصاویر دارند، و دوم اینکه مواردی در تصاویر پزشکی وجود دارد که داده های مربوط به دو اصطلاح یا برچسب مختلف، تفاوت های بسیار کمی دارند و عملا مرز مشخص بین کلاس ها وجود ندارد. در هر دو این موارد دسته بندی کننده مبتنی بر نزدیک ترین همسایه انتخاب مناسب تری می باشد [۳۵].

۴-۲ ارزیابی و تحلیل نتایج

به منظور بررسی دقت سیستم پیشنهادی، ارزیابی را به دو روش می توان انجام داد: روش اول (دوبعدی_سطح-اسلایس)، روی اسلایس ها یا برش های تصویر سی تی به صورت جداگانه انجام می شود، یعنی هر برش را به عنوان یک تصویر در نظر گرفته و پاسخ سوالات را با استفاده از آن تصویر به دست آورد. روش دوم (سه بعدی_سطح-حجم)، برای هر تصویر سی تی سه بعدی (حجمی)، پاسخ هر سوال را به دست آورد، بدین معنا که برای هر بیمار با استفاده از بررسی تمامی برش ها پاسخ سوالات را به دست آورد. برای بررسی دقت در هر دو روش، در این مقاله فقط برش های کلیدی هر بیمار که شامل بهترین نمایش از کبد و ضایعه ها

¹ Binary crossentropy

² categorical_crossentropy

³ Batch learning

⁴ <https://colab.research.google.com>

⁵ Ensemble

⁶ Random subspace

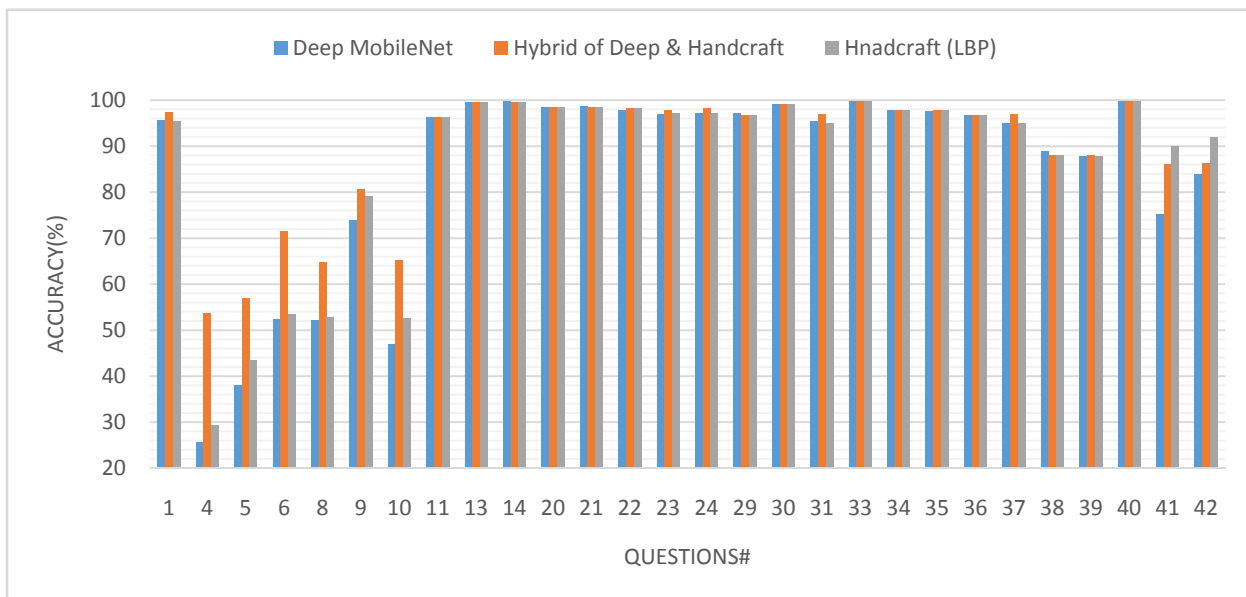
است که نقطه‌ی مشترک روش پیشنهادی و مرجع [۱۳] که دقت-های بالاتری را نسبت به روش‌های مراجع دیگر در جدول ۲ به دست آورده‌اند، تاکید بر کشف ویژگی‌های متناسب برای پاسخ-دهی به هر سوال به صورت جداگانه و عدم استفاده از مجموعه ویژگی مشترک برای تمامی سوالات می‌باشد.

در مقایسه نتایج در حوزه‌ی دوبعدی و سه‌بعدی، دلیل بالاتر بودن دقت در حوزه‌ی دوبعدی را می‌توان مشابه بودن برش‌ها در مجموعه‌های آموزش و اعتبارسنجی دانست، زیرا که ممکن است متعلق به یک بیمار باشند. که در مسائلی که دیتاست‌های با حجم بالا در اختیار باشد و بیماران با شرایط مشابه با بیمار ورودی وجود داشته باشد، بیشتر اتفاق می‌افتد. اما در مسئله‌ی سه‌بعدی برش‌های متعلق به هر بیمار با هم و در کنار هم و به‌عنوان یک نمونه‌ی مجزا، بررسی می‌شوند.

کارآمد و اختصاصی برای پاسخ‌دهی به همان سوال است، برش‌مرد. زیرا ماهیت سوالات در هر گزارش رادیولوژی با یکدیگر متفاوت است و قطعا یک مجموعه ویژگی استخراج‌شده برای پاسخ‌دهی به تمامی سوالات مناسب نیست، درحالی‌که کارهای مشابه برای تمامی سوالات از بردارهای ویژگی یکسان استفاده کرده‌اند. مزیت روش پیشنهادی نسبت به مرجع [۱۳] استفاده از روش‌های یادگیری عمیق است، اگرچه حدود ۱,۷% دقت پایین-تری به دست آورده است، اما فرآیند یافتن بهترین ویژگی برای هر سوال با استفاده از شبکه‌ی عمیق مختص همان سوال انجام شده است که کاملا اتوماتیک است و نیاز به بررسی جداگانه‌ی مجموعه ویژگی‌های متفاوت برای هر سوال وجود ندارد. از طرف دیگر روش پیشنهادی فقط یک نوع ویژگی دستی استفاده می‌کند که این خود در مقایسه با روش‌های دیگر هیبریدی [۲۳] که از چندین نوع ویژگی استفاده می‌کنند حجم محاسبات کمتری دارد. لازم به ذکر

جدول ۱: میانگین دقت و انحراف هر روش برای تمامی سوالات روی تصاویر دوبعدی (سطح-اسلایس)

روش	بعد بردار ویژگی	دسته‌بندی‌کننده	دقت (%)	انحراف معیار
شبکه عصبی عمیق موبایل نت انتها-به-انتها	-	-	۹۰,۳	۱,۷۲
مبتنی بر ویژگی دستی (الگوی دودویی محلی)	۷۲	RSKNN	۹۱,۶۷	۰,۰۶
مبتنی بر ویژگی عمیق (موبایل نت)	۱۰۰	RSKNN	۹۱,۶۹	۰,۰۸
روش پیشنهادی هیبریدی (دستی و عمیق)	۷۲+۱۰۰=۱۷۲	RSKNN	۹۳,۱۵	۰,۰۵



شکل ۳ نمودار دقت روش‌های ارائه‌شده روی تصاویر دوبعدی به تفکیک سوالات

جدول ۲: مقایسه نتایج روش پیشنهادی روی تصاویر سه بعدی (سطح-حجم) با مقالات پیاده سازی شده روی دیتاست ImageCLEF Liver CT2015

روش	نوع ویژگی	بعد	دسته بندی کننده	دقت (%)
مرجع [۹]	SIFT	۱۰۰۰	Weighted Nearest Neighbour	۸۸,۷
مرجع [۱۰]	GLCM + 3D Gabor	۱۱۱	Random Forest	۸۴
مرجع [۱۱]	Gabor of BIMFs	۹۶۰	CBIR/majority voting	۸۸,۹
مرجع [۱۳]	DLBP/BoW/Shape	۵۰۰-۳۶	SVM و RSKNN	۹۳,۱
روش پیشنهادی هیبریدی	Deep(MobileNet)+LBP	۱۷۲	RSKNN	۹۱,۴

از تصاویر پزشکی، شبکه های عصبی پیچشی موبایل نت که بستر سخت افزاری سبک تر و پارامترهای کمتری دارند (در مقایسه با سایر شبکه های عمیق) انتخاب شده اند زیرا این ویژگی ها در مصارف پزشکی که سخت افزار و داده های آموزشی محدود است؛ قابل توجه و مهم هستند، و همچنین از ویژگی بافت الگوی دودویی محلی که در دسته بندی تصاویر پزشکی و غیرپزشکی پرکاربرد می باشد، استفاده شده است. برای کاهش تاثیر منفی کم بودن داده های آموزشی و پیشگیری از بیش برآزش شبکه های عمیق از روش های یادگیری انتقال، افزایش داده و حذف تصادفی بهره برده ایم. برای دسته بندی تصاویر با توجه به اینکه توزیع داده ها برای سوالات مختلف تفاوت های بسیار زیادی دارند روش ترکیبی مبتنی بر نزدیک ترین همسایه انتخاب شده است. روش پیشنهادی روی تولید گزارش ساخت یافته در مورد تصاویر کبد از ImgeCLEF Liver ct Annotation مورد ارزیابی قرار گرفت. نتایج نشان می دهد که روش پیشنهادی هیبریدی دقت بالاتری نسبت به سایر روش ها دارد.

در پژوهش های آتی می توان امکان استفاده از روش پیشنهادی در دیگر حوزه های پزشکی را مورد بررسی قرار داد، زیرا انتظار می رود روش در دیگر حوزه ها نیز موفق عمل کند. به این دلیل که برخلاف روش های کلاسیک که انتخاب روش استخراج ویژگی توسط توسعه دهندگان نرم افزار بر اساس محتوای تصویر یا ارگان یا سوال پرسیده شده، می بایست انجام شود، در این روش بخش عمده ای از ویژگی ها توسط شبکه عمیق و متناسب با تصویر و سوال استخراج می شود. همچنین برای افزایش دقت سیستم می توان شبکه های عمیق متنوع تری را آزمود. از طرفی استفاده از روش های پیش پردازش و قطعه بندی برای برخی سوالات مرتبط با شکل، در مراحل پیش از استخراج ویژگی، پیشنهاد می شود.

عملکرد شبکه موبایل نت در مسئله ای ما که تعداد داده های آموزشی کم است (۵۰ بیمار) با استفاده از روش های افزایش داده و یادگیری انتقال تقویت شده است. اما نتایج بیانگر آن است که اگرچه شبکه موبایل نت یک شبکه کم پارامتر و کم حجم است اما همچنان نیاز به داده های آموزشی با حجم بالا وجود دارد. بنابراین همانطور که نتایج نشان داده و پیش از این اشاره شده، یکی از مطمئن ترین راه حل ها در پاسخ دهی به سوالات پزشکی که تعداد داده های آموزشی کم و سوالات متنوع و گسترده هستند، بهره بردن از ترکیب ویژگی های استخراج شده از شبکه ای عمیق و دستی می باشد.

همانطور که در شکل ۳ و ۴ مشخص شده است روش های بررسی شده برای سوال ۴ که در مورد شکل ضایعات کبدی (Lesion hasAreaMarginType) و چهارکلاسی (ill defined, irregular, lobular, well deifined) می باشد، دقت پایینی دارند، که این مسئله یکی از چالش های باقیمانده در پاسخ دهی به سوالات می باشد. بررسی ها نشان می دهد که دلیل اصلی پایین بودن دقت در این چنین سوالات کم بودن نمونه های آموزشی برای برخی از کلاس ها، پیچیدگی دسته بندی و وابستگی به عملیات قطعه بندی تصویر می باشد که با بهبود شرایط مجموعه آموزشی یا اجرای برخی روش های پیش پردازش در پاسخ دهی به برخی سوالات، انتظار بهبود دقت سیستم را داریم.

۵ نتیجه گیری و پیشنهاد برای پژوهش های آتی

در راستای حل چالش عدم وجود اطلاعات تخصصی در مورد تصاویر پزشکی در این پژوهش به ارائه ی سیستم تولید خودکار گزارش ساخت یافته رادیولوژی در مورد تصاویر پزشکی پرداخته شده، که قادر است به سوالات ساخت یافته در مورد محتوای تصاویر پاسخ دهد. در سیستم پیشنهادی به منظور استخراج ویژگی



شکل ۴ نمودار دقت و زیان تعدادی از سوالات (شماره و تعداد گزینه در عنوان هر نمودار مشخص شده) با استفاده از شبکه موبایل نت انتها-به-انتها

[2] Sugimori, H. Classification of computed tomography images in different slice positions using deep learning. *Journal of healthcare engineering* 2018, 2018.

[3] Alkhawlan, M.; Elmogy, M.; El Bakry, H. Text-based, content-based, and semantic-based image retrievals: A survey. *Int. J. Comput. Inf. Technol* 2015, 4.

مراجع

[1] Tommasi, T.; Orabona, F.; Caputo, B. Discriminative cue integration for medical image annotation. *Pattern Recognition Letters* 2008, 29, 1996–2002.

- [16] Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 2014.
- [17] He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
- [18] Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2818–2826.
- [19] Chollet, F. Xception: Deep learning with depthwise separable convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1251–1258.
- [20] Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 2017.
- [21] Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-resnet and the impact of residual connections on learning. Thirty-First AAAI Conference on Artificial Intelligence, 2017.
- [22] Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012, pp. 1097–1105.
- [23] Zhang, J.; Xia, Y.; Xie, Y.; Fulham, M.; Feng, D.D. Classification of medical images in the biomedical literature by jointly using deep and handcrafted visual features. IEEE journal of biomedical and health informatics 2017, 22, 1521–1530.
- [24] Jia, Y.; Shelhamer, E.; Donahue, J.; Karayev, S.; Long, J.; Girshick, R.; Guadarrama, S.; Darrell, T. Caffe: Convolutional architecture for fast feature embedding. Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014, pp. 675–678.
- [25] Lee, S.L.; Zare, M.R.; Muller, H. Late fusion of deep learning and handcrafted visual features for biomedical image modality classification. IET Image Processing 2018, 13, 382–391.
- [26] Marvasti, N.B.; Garcia, M.d.M.R.; Uskudarli, S.; Montes, J.F.A.; Acar, B. Overview of the ImageCLEF 2015 liver CT annotation task. CLEF (Working Notes), 2015.
- [27] Tanner, M.A.; Wong, W.H. The calculation of posterior distributions by data augmentation. Journal of the American statistical Association 1987, 82, 528–540.
- [28] Hussain, Z.; Gimenez, F.; Yi, D.; Rubin, D. Differential data augmentation techniques for medical imaging classification tasks. AMIA Annual Symposium
- [4] Qayyum, A.; Anwar, S.M.; Awais, M.; Majid, M. Medical image retrieval using deep convolutional neural network. Neurocomputing 2017, 266, 8–20.
- [5] Afshar, P.; Mohammadi, A.; Plataniotis, K.N.; Oikonomou, A.; Benali, H. From handcrafted to deep-learning-based cancer radiomics: Challenges and opportunities. IEEE Signal Processing Magazine 2019, 36, 132–160.
- [6] Singh, S.; Ho-Shon, K.; Karimi, S.; Hamey, L. Modality Classification and Concept Detection in Medical Images Using Deep Transfer Learning. 2018 International Conference on Image and Vision Computing New Zealand (IVCNZ). IEEE, 2018, pp. 1–9.
- [7] Liu, M.; Zhang, J.; Nie, D.; Yap, P.T.; Shen, D. Anatomical landmark based deep feature representation for MR images in brain disease diagnosis. IEEE journal of biomedical and health informatics 2018, 22, 1476–1485.
- [8] Yasar, A.; Saritas, I.; Korkmaz, H. Computer-Aided Diagnosis System for Detection of Stomach Cancer with Image Processing Techniques. Journal of medical systems 2019, 43, 99.
- [9] Kumar, A.; Dyer, S.; Kim, J.; Li, C.; Leong, P.H.; Fulham, M.; Feng, D. Adapting content-based image retrieval techniques for the semantic annotation of medical images. Computerized Medical Imaging and Graphics 2016, 49, 37–45.
- [10] Nedjar, I.; Mahmoudi, S.; Chikh, M.A.; Abi-Yad, K.; Bouafia, Z. Automatic Annotation of Liver CT Image: ImageCLEFmed 2015. CLEF (Working Notes), 2015.
- [11] Nedjar, I.; Mahmoudi, S.; Chikh, M.A. Content-based Medical Image Retrieval for Liver CT Annotation. Transactions on Machine Learning and Artificial Intelligence 2017, 5.
- [12] Spanier, A.B.; Caplan, N.; Sosna, J.; Acar, B.; Joskowicz, L. A fully automatic end-to-end method for content-based image retrieval of CT scans with similar liver lesion annotations. International journal of computer assisted radiology and surgery 2018, 13, 165–174.
- [13] Loveymi S, Dezfoulian MH, Mansoorizadeh M. Generate Structured Radiology Report from CT Images Using Image Annotation Techniques: Preliminary Results with Liver CT. Journal of digital imaging. 2019 Nov 14:1–6.
- [14] Baltruschat, I.M.; Nickisch, H.; Grass, M.; Knopp, T.; Saalbach, A. Comparison of deep learning approaches for multi-label chest X-ray classification. Scientific reports 2019, 9, 6381.
- [15] Li, Y.; Huang, H.; Xie, Q.; Yao, L.; Chen, Q. Research on a surface defect detection algorithm based on MobileNet-SSD. Applied Sciences 2018, 8, 1678. Version September 8, 2019



سمیرا لویمی مدرک کارشناسی (مهندسی کامپیوتر نرم‌افزار) و کارشناسی ارشد (مهندسی کامپیوتر هوش مصنوعی) خود را از دانشگاه شهید چمران اهواز در سال‌های ۱۳۸۶ و ۱۳۸۸ دریافت نمود. ایشان هم‌اکنون در مقطع دکتری تخصصی در رشته مهندسی کامپیوتر در دانشگاه بوعلی‌سینا همدان مشغول به تحصیل می‌باشد. زمینه‌های تحقیقاتی وی بینایی ماشین، پردازش تصاویر پزشکی، سیستم‌های پرسش و پاسخ بصری و یادگیری عمیق است.



میرحسین دزفولیان مدرک کارشناسی و کارشناسی ارشد علوم کامپیوتر از دانشگاه صنعتی شریف در سال‌های ۱۳۶۳ و ۱۳۶۶ دریافت و دکتری با تخصص تشخیص الگو از دانشگاه ولنگنگ، استرالیا در سال ۱۳۷۴ اخذ نموده‌اند. ایشان استادیار همکار گروه کامپیوتر دانشگاه بوعلی‌سینا هستند و زمینه‌های مورد علاقه پژوهشی وی سیستم‌های خبره، سیستم‌های پرسش و پاسخ، بازیابی و پردازش متون و تصاویر می‌باشد.



محرم منصوریزاده کارشناسی خود را در رشته مهندسی کامپیوتر در سال ۱۳۸۰ از دانشگاه اصفهان و کارشناسی ارشد و دکتری خود را در همین رشته در سال‌های ۱۳۸۲ و ۱۳۸۸ از دانشگاه تربیت مدرس اخذ نموده است. او استادیار گروه مهندسی کامپیوتر دانشگاه بوعلی‌سینا است و به پژوهش در زمینه‌های یادگیری ماشین، ارتباط هوشمند انسان و ماشین، پردازش تصویر، ویدیو و صدا و همچنین زبان‌شناسی رایانشی مشغول است.

Proceedings. American Medical Informatics Association, 2017, Vol. 2017, p. 979.

- [29] Hollemans, M. MobileNet version 2. Machinethink.net. <http://machinethink.net/blog/mobilenet-v2/>. 2018.
- [30] Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510–4520.
- [31] Baldi, P.; Sadowski, P.J. Understanding dropout. Advances in neural information processing systems, 2013, pp. 2814–2822.
- [32] Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research 2014, 15, 1929–1958.
- [33] Warde-Farley, D.; Goodfellow, I.J.; Courville, A.; Bengio, Y. An empirical analysis of dropout in piecewise linear networks. arXiv preprint arXiv:1312.6197 2013.
- [۳۴] امیرفتحیان، امیررضا، ابراهیم نژاد، حسین. (۱۳۹۸). تولید طرح چهره از تصویر نمونه با کدگذاری الگوی دودویی محلی. مجله ماشین بینایی و پردازش تصویر. (آماده انتشار).
- [35] Kuncheva, L.I. Combining pattern classifiers: methods and algorithms; John Wiley & Sons, 2014.