

چکیده:

سرطان سینه یک بیماری مهلك و در عین حال قابل درمان در ابتدای تشخیص است در گذشته برای تشخیص سرطان به صورت دستی و توسط پزشک انجام میشده است که با خطاهایی بالایی به خصوص در تشخیص سرطان سینه در حالات اولیه این بیماری بود با ورود هوش مصنوعی به این بیماری دقت بالایی در تشخیص بیماری حاصل شد اما مشکل بزرگتر در تشخیص نابه جای این بیماری بود که حتی با دقت کم می تواند خسارات جبران ناپذیری را به بیمار وارد کند ما در این مقاله با سه الگوریتم تشخیص سرطان آشنا می شویم و راهکار هایی را برای بهبود آن ارائه میدهم الگوریتم KNN که ساده ترین الگوریتم تشخیص سرطان سینه تا الگوریتم هایی دقیقی مثل CNN و SVM که دارای دقت بالایی در تشخیص این بیماری اند

کلید واژه : سرطان سینه , svm , cnn , knn , ماموگرافی , hough , استخراج ویژگی

مقدمه :

سرطان یک چالش مهم در سلامت عمومی برای دنیای امروز است. طبق گزارش IARC¹ (آژانس بین المللی تحقیقات سرطان) و WHO² (سازمان بهداشت جهانی) 8/2 میلیون مرگ ناشی از سرطان در سال 2012 ثبت شده است و انتظار میرود پیش از سال 2030 ، 27 میلیون مورد جدید از این بیماری را شاهد باشیم [1,6]. ، سرطان سینه³ یکی از مشکلات بهداشتی قابل توجه در جهان است. [2] به ویژه شایعترین نوع بیماری سرطانی در بین زنان جهان غرب محسوب می شود . تقریباً یک دهم زنان در طول زندگی خود از آن رنج می برند ، که نیمی از آنها زنده نمی مانند. [3,4] در ایالات متحده ، از هر هشت زن یک نفر به سرطان سینه مبتلا می شوند. هر ساله حدود 40000 نفر به دلیل سرطان جان خود را از دست می دهند. حدود 1.3 میلیون نفر وجود دارند که به دلیل سرطان تحت جلسات درمانی قرار گرفته اند. در آسیا ، درصد مبتلایان به سرطان سینه 1.7 درصد در سال در حال افزایش است [5,7]

چگونه سرطان ایجاد میشود ؟؟

تومور نوعی بیماری ناخوشایند است و ویژگی های اصلی بیماری سرطان سینه رشد بیش از حد سلول ها در یک ناحیه خاص از بدن است . این رشد سلولی نیز به نوعی به عنوان تومور شناخته شده است. سرطان هنگامی ایجاد می شود که بیماری از بافت های تنیده سرطان در بافت های سینه انسان ایجاد می شود [1,2,14].

زمان پیدایش پردازش تصویر در پزشکی:

از زمان استفاده از هوش مصنوعی (AI) در رایانه ، مطالعات در زمینه پردازش تصویر پزشکی آغاز شده است. اولین تحقیق در پردازش تصویر پزشکی در دهه 1960 ظاهر شد و روشهای مختلفی در تجزیه و تحلیل تصاویر رادیوگرافی مورد بحث قرار گرفت. [5] پس از آن ، در دهه 1970 ، تحقیقات دیگری پدید آمد که سعی در توصیف تکنیک هایی دارد که در کشف رادیوگرافی غیر طبیعی قفسه سینه تخصص دارند. این مطالعات در [8] ذکر شده است. استفان جاگر⁴ و الکساندرس کا آرگریس⁵ با استفاده از فناوری پردازش تصویر پزشکی و با استفاده از تصاویر رادیولوژیک قفسه سینه ، تشخیص هایی را در مورد کشف سل انجام دادند. [5]

در دهه گذشته ، فناوری پردازش تصویر به یک زمینه جالب در پزشکی تبدیل شده است. زیرا پردازش تصویر در مراحل ابتدایی نقش مهمی در کشف بیماری ها دارد و این امر باعث تسهیل در درمان این بیماری ها می شود. [5] طبقه بندی تومور یک کار مهم در

¹ Inter National Agency for Research on Cancer

² World Health Organization

³ breast-cancer

⁴ Stefan Jaeger

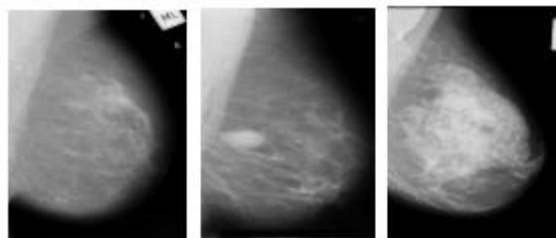
⁵ Alexandros Ka Argyris

تشخیص پزشکی محسوب میشود. روشهای محاسبات نرم افزاری به دلیل عملکرد طبقه بندی آنها در تشخیص بیماریهای پزشکی اهمیت زیادی دارند. تشخیص و طبقه بندی تصاویر پزشکی یک کار چالش برانگیز است.[3] اگر ناهنجاری های سرطان سینه زود یا در همان اول پیش بینی و تشخیص داده شود ، بیشترین احتمال بهبودی وجود دارد.[2] در حوزه های پزشکی ، حجم و پیچیدگی داده های جمع آوری شده با سرعت زیادی در حال رشد است. این اطلاعات علاوه بر اطلاعات ناشی از مطالعات بالینی ، شامل داده های دیگر در مورد بیماران است .[9,3] تجزیه و تحلیل چنین داده هایی امکان ایجاد فرضیه های پزشکی جدید یا تأیید فرضیه های موجود را ممکن می سازد. این تا حدی بر مطالعات پزشکی سنتی که به تعداد کمی از پارامترها و تعداد کمی از موارد محدود شده بود غلبه می کند. ارزشهای اجتماعی و اقتصادی تشخیص سرطان سینه بسیار زیاد است [10,3]. در نتیجه ، در این زمینه اخیراً نظر بسیاری از محققان در زمینه داده کاوی را به خود جلب کرده است [11,12,3]. تلاش های آنها رویکردهای مختلفی را برای تشخیص خودکار این بیماری ایجاد کرده است تشخیص سرطان سینه میتواند با روش های تصویربرداری مختلفی مانند ماموگرافی، ترموگرافی و تصویربرداری رزونانس مغناطیسی انجام شود(1,13) در این مقاله ما سه الگوریتم svm و knn و cnn را بر روی مجموعه داده های ماموگرافی با هم مقایسه خواهیم کرد ابتدا خود الگوریتم و سپس راه هایی برای بهبودی آن نیز مورد مطالعه قرار خواهیم داد .

ساختار این مقاله به شرح زیر است :بخش اول مقدمه ایی بر سرطان سینه و همچنین تاریخچه ایی از آن را شامل میشود بخش دوم کار بر روی مجموعه داده را معرفی میکند، بخش سوم پیاده سازی انواع روش بر روی داده ها را مورد بحث قرار میدهد، بخش چهارم مقایسه الگوریتم ها با هم را بیان میکند و سرانجام در بخش پنجم با نتیجه گیری این کار به پایان می رسد.

1.1 تهیه تصویر:

تصاویر ماموگرافی از نوع طبیعی ، خوش خیم و بدخیم برای انواع غده چربی سینه که از پایگاه داده تجزیه و تحلیل تصویر ماموگرافی بانک اطلاعاتی MIAS جمع آوری شده است. این تصاویر ماموگرافی دیجیتالی دارای مرزهای 50 میلی متر پیکسل است. تصاویر کاهش⁶ و کلیپ شده⁷ است به طوری که هر تصویر 1024*1024 پیکسل است وبا فرمت PGM می باشد. (15,1,2)



الف: خوش خیم ب:نرمال پ:بدخیم

اما این پایگاه داده ممکن است دارای نویز ها و اطلاعات ناخواسته ایی باشد که به اشتباه علامت گذاری شده باشند ما با انجام یک مرحله پیش پردازش میتوانیم این نویز ها را به حداقل برسانیم ودقت های بالاتری را در الگوریتم های خود به دست آوریم پس بر روی داده ها یک مرحله پیش پردازش انجام می دهیم .

1.2 پیش پردازش:

برای مثال شکل 1 ناحیه نامطلوب سرطان سینه و تومور را نشان می دهد.منطقه ای که با رنگ زرد مشخص شده مربوط به بخشی از سینه است ، برچسب و پس زمینه نامطلوب با استفاده از دایره آبی و دایره قرمز نشان داده شده است - که به اشتباه به عنوان تومور شناخته شده اند ابتدا برچسب نامطلوب با استفاده از روش

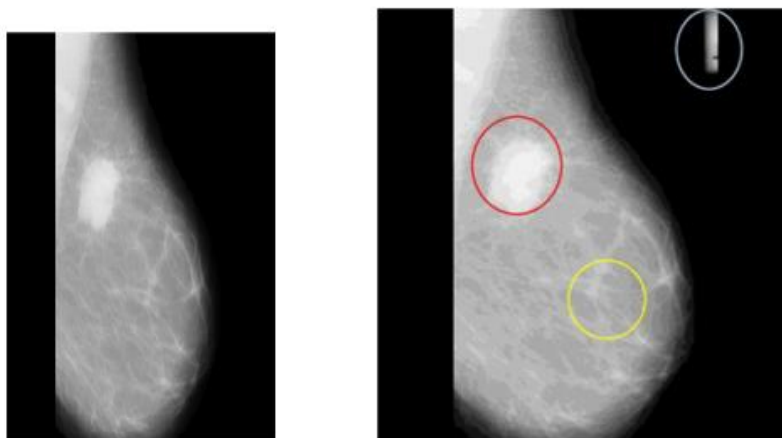
⁶ reduced

⁷ clipped

گرادیان مبتنی بر آستانه⁸ حذف می شود. چندین عملکرد مورفولوژیکی برای پنهان کردن برجسب نامناسب انجام می شود. پس از حذف کردن برجسب ها مهمترین کار پر کردن محل نویز هاست که با استفاده از فرمول ، نقشه مرز باینری تصویر که در آن از روش گرادیان مبتنی بر آستانه استفاده شده است استفاده می کنیم .

$$f(x,y) = \begin{cases} 1 & \text{if } G(x,y) \geq GT \\ 0 & \text{else} \end{cases} \quad (1)$$

: فرمول نقشه بندی مرز



حذف برجسب ها از تصویر

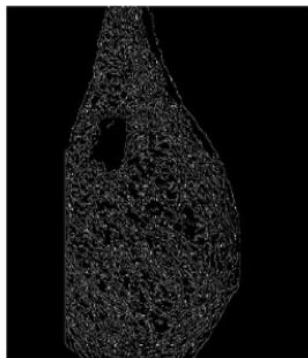
شکل 1

در اینجا GT یک گرادیان آستانه است که با استفاده از روش تطبیقی Otsu محاسبه می شود . پس از این تصویر باینری با استفاده از اساس ساختار الماسی (شکل 2) گسترش می یابد و یک ماسک را ایجاد میکند اکنون این ماسک با تصویر اصلی ضرب می شود.

ما برای به دست آوردن ویژگی های موثر از روش مرز بندی برای نشان دادن بافت های سینه استفاده میکنیم طبق مقاله 2 مرز بندی تصویر میتواند به دو صورت هوشمند (شکل 3) و غیر هوشمند (شکل 4) باشد که ما فقط از روش مرز بندی هوشمند استفاده می کنیم. خوب حالا برای به دست آوردن ویژگی ها ما باید بافت های سینه را مورد بررسی قرار دهیم بنابراین این ماهیچه ها باید برای استخراج ویژگی های موثر با استفاده از روش حداکثر احتمال منطقه ای که حداکثر احتمال وجود بافت های سرطان در آن است را شناسایی کرده و ویژگی های آن را استخراج می کند معیار حداکثر احتمال که بافت های سرطان در آن است یا نه در زیر فرموله بندی شده است [16,2].

$$L\theta = \ln(P(X|\theta))$$

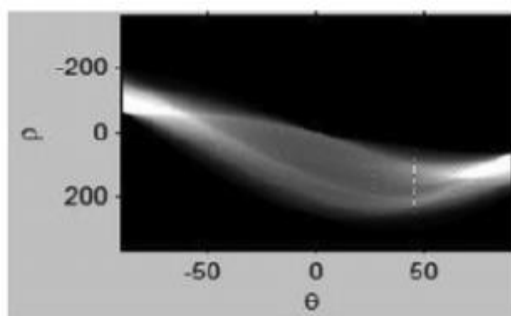
⁸ gradient based threshold



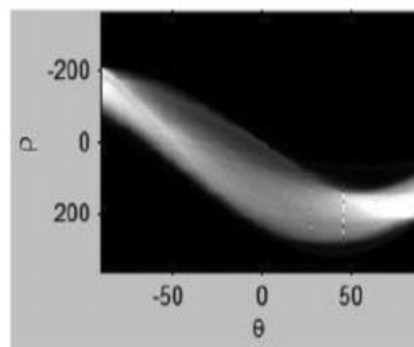
مرز بندی تصویر



شکل 2



مرز بندی غیر هوشمند (شکل 4)



مرز بندی هوشمند (شکل 3)

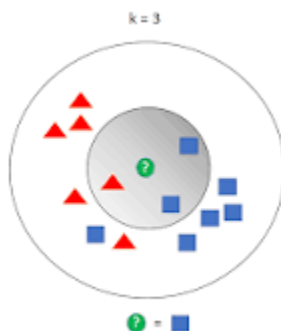
1 بهبود روش k نزدیک ترین همسایه یا KNN :

الگوریتم نزدیکترین همسایه ها (KNN) یکی از ساده ترین الگوریتم هایی است که از آن به طور گسترده در تجزیه و تحلیل پیش بینی ها استفاده می شود.

برای شناخت این الگوریتم ابتدا مثالی میزنیم فرض کنید که شما فروشنده اید و دو دسته مشتری دارید. مشتری های مرفهی که بالای 45 سال سن دارند و وقتی به فروشگاه شما می آیند و بیشتر از 100 هزار تومان خرج میکنند . و مشتری های عادی که زیر 25 سال دارند و کمتر از 10 هزار تومان خرج میکنند حال فکر کنید که مشتری بالای 45 سال به شما مراجعه کند آنوقت آن شخص جز کدام دسته خواهد بود ؟

خوب جواب این سوال را میتوان بر اساس داده های قبلی به دست آورد در الگوریتم های یادگیری ماشین نیز بر همین اساس با استفاده از داده های آموزش (قبلی) برای داده های تست نتیجه گیری می شود .

الگوریتم knn بر اساس داده های قبلی ابتدا k همسایه نزدیک به داده جدید را پیدا می کند (برای پیدا کردن همسایه های نزدیک از فاصله اقلیدسی استفاده میکند که در مقاله 22 توضیح داده شده است) و بعد بین آن ها رای گیری انجام می دهد هر دسته که رای بیشتری داشته باشد داده جدید به آن دسته تعلق میگیرد.



بر اساس شکل بالا میتوان نتیجه گرفت داده جدید ما (؟) متعلق به دسته مربع میباشد k همیشه مقداری فرد است تا نتیجه رای گیری هرگز به تساوی نینجامد.

برای بهینه سازی عملکرد و سرعت بخشیدن به فرایند الگوریتم KNN، در این مقاله راه حل جدیدی برای سرعت بخشیدن به الگوریتم KNN بر اساس خوشه بندی و ویژگی های فیلتر ارائه شده است. در اینجا همچنین شامل پیشرفت دیگری بر اساس ضرایب قابلیت اطمینان است که طبقه بندی⁹ دقیق تری را تضمین می کند نیز می شود.

1.1 رویکرد پیشنهادی:

برای بهبود این الگوریتم ابتدا باید محدودیت های آن را بشناسیم

. محدودیت K-nearest neighbors:

محدودیت های الگوریتم KNN به شرح زیر است:

- 1 الگوریتم KNN کند است زیرا هر بار همه موارد را مرور می کند
- 2 این الگوریتم در برابر ابعاد آسیب پذیر است
- 3 این الگوریتم به صفات نامرتبط و مرتبط حساس است.
- 4 انتخاب اشتباه فاصله یا مقدار k عملکرد را خراب می کند.

خوب برای حذف چنین محدودیت هایی الگوریتم KNN در تلاش برای غلبه بر محدودیت های خود، پیشرفت های زیادی را مشاهده کرده است: [17] که می توان به اضافه کردن وزن به KNN و ضریب اطمینان و الگوریتم خوشه بندی اشاره کرد

برای حل مشکل 2 و 3:

بیلی برای ارائه همسایه وزن را به KNN کلاسیک اضافه کرد، و¹⁰ (WKNN) را معرفی کرد [18]. اما وزن چه تاثیری در الگوریتم ما دارد؟ خوب برای پاسخ به این سوال باید با ذات knn آشنا شویم در واقع، در KNN، نزدیکترین همسایه ها بدون در نظر گرفتن میزان شباهت آنها با نمونه جدید در پیش بینی به مقدار یکسانی تأثیر می گذارند. ولی در WKNN براساس این عقیده است که نمونه هایی که خصوصاً به نمونه جدید نزدیک هستند باید نسبت به نمونه هایی که از آن دورتر هستند، وزن بیشتری در نظر بگیرند، یونگ¹¹، و همکارانش [24] از الگوی جدید KNN، که با خوشه بندی توسعه داده شده، برای طبقه بندی متن استفاده کرده است

. خوب حتماً میگوید چرا از خوشه بندی استفاده کرده؟

⁹ classification

¹⁰ weighted K-nearest neighbors

¹¹ Yong

الگوریتم خوشه بندی داده هایی را که دارای ویژگی ها نزدیکی به هم اند را در یک دسته قرار میدهد و آن ها را به شکل یک دسته در می آورد که باعث می شود مشکل 1 ما در knn به راحتی حذف شود این الگوریتم داده را با مرکز خوشه ها مقایسه می شود و پی میبرد که به آن دسته تعلق دارد یا نه (تعداد خوشه به عهده خود برنامه نویس است) یکی از الگوریتم های معروف خوشه بندی K-MEAN است که ما نیز از آن بهره گرفته ایم بنابراین ، طبقه بندی نمونه جدید در یکی از خوشه های k به جای مقایسه آن با N نمونه های اولیه ، زمان محاسبه الگوریتم را به k / n کاهش می دهد

نرمال سازی¹³: تنها مشکل 4 باقی می ماند که ما برای حل این مشکل از نرمال سازی استفاده می کنیم که فواصل را به صورت استاندارد در فواصل 0 و 1 در می آورد که باعث بدست آوردن دقت بهتر میشود که برای این کار نیاز داریم که از فرمول نرمال سازی زیر استفاده کنیم

$$e_{ij} \leftarrow \frac{e_{ij} - \min_j(e_{ij})}{\max_j(e_{ij}) - \min_j(e_{ij})}$$

فرمول نرمال سازی

اندازه گیری ضریب اطمینان¹⁴ : ضرایب قابلیت اطمینان باعث می شود از یک طرف انتخاب صفات قابل توجه و مهم تصویر و از سوی دیگر کاهش ابعاد را برای ما مقدور می سازد.

معایب این الگوریتم تقریباً تمام معایب الگوریتم knn را بر طرف کرده است به جز

1: ذخیره تمامی داده های آموزشی

2: حساس بودن به پارامتر k

3: دقت پایین تر نسبت به دو الگوریتم دیگر

4: این الگوریتم فقط برای طبقه بندی استفاده می شود

مزایا:

1: حساس نبودن به مقدار های مثل فاصله با انجام نرمال سازی

2: سرعت بیشتر نسب به الگوریتم knn معمولی به خاطر به کار بردن الگوریتم kmeans

3: سرعت و ساده تر بودن نسبت به سه روش پیشنهادی دیگر

4: نسبت به سه روش knn معمولی و cnn معمولی (ANN) و svm معمولی دارای دقت بالاتری است

5: قابلیت ادغام با دو روش دیگر را دارد

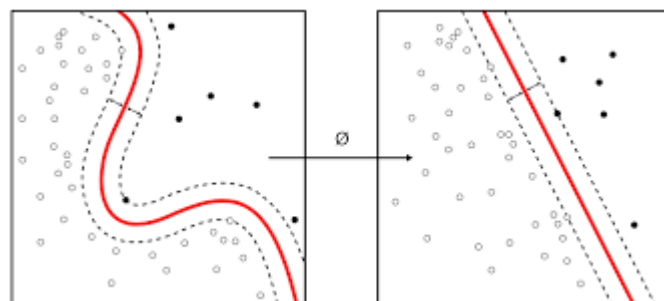
2: بهبود روش ماشین بردار پشتیبان یا svm :

¹² Clustering

¹³ Normalization

¹⁴ Reliability measure

استدلال ماشین بردار پشتیبان این است که یک مرز را برای انتخاب داده جدید که در کدام دسته قرار بگیرد مشخص میکند مثل یک خط یا یک مسیر غیر خطی که دو دسته را از هم جدا می کند.



SVM غیر خطی

SVM خطی

داده ها براساس این مرز ها دسته بندی میشوند و داده های جدید بر اساس ویژگی ها خود که به یک سمت مرز نزدیکتر است در یک سمت مرز قرار می گیرند.

روش svm دارای دقتی بین 85.9 است که با استفاده از روش که ما ارائه می دهیم تا دقت 94 در صد هم پیش خواهد رفت. (برگرفته از مقالات 1 و 2) برای این کار ما به یک تبدیل گر Hough نیاز داریم. تبدیل Hough بر روی ROI (منطقه ای که احتمال وجود تومر آن جاست) انجام می شود تا ویژگی ها را استخراج کند. یک روش موثر برای استخراج ویژگی و تشخیص شکل خاص در تصویر است. همچنین تصویر را به صورت دوبعدی در می آورد که این کار باعث سرعت در تشخیص تصویر می شود و در نتیجه در طبقه بندی می شود تبدیل Hough باعث تحمل در برابر نویز شکاف ها و انسداد در ماموگرافی ها می شود. که فرمول آن در [2] آمده است خوب پس از استخراج ویژگی با hough به جای classifier از svm استفاده می کنیم طبقه بندی SVM برای ماموگرافی با استفاده از فرمول زیر بدست می آید

$$f_{svm}(X) = W_t \varphi(X) + b \quad (9)$$

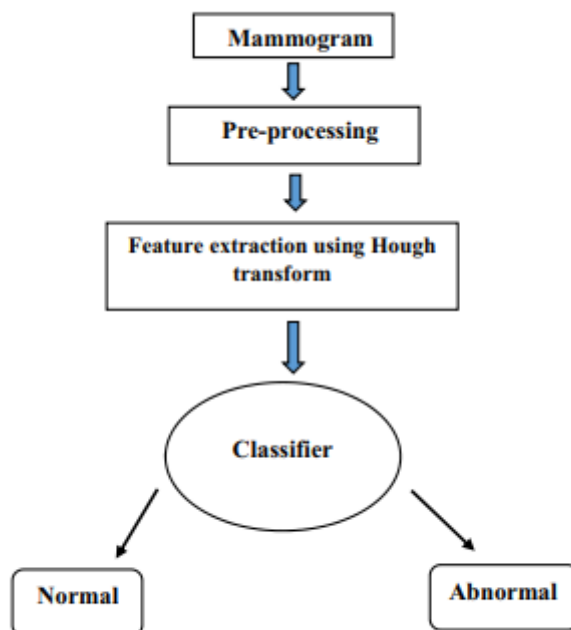


Fig. 1. Proposed method.

شکلی کلی از مراحل svm پیشنهادی (شکل 5)

مزایا [20] :

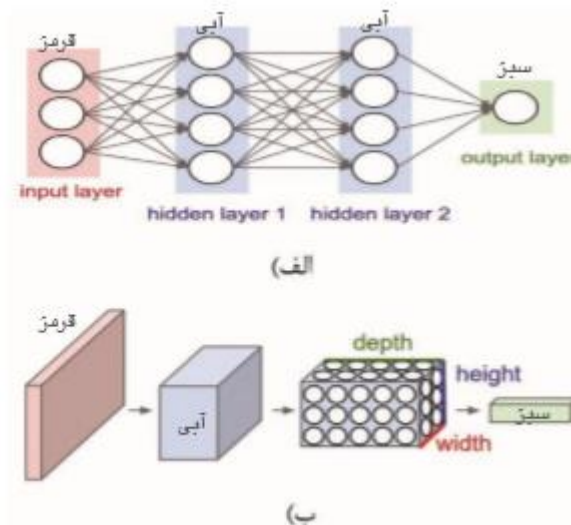
- 1: پیچیدگی کمتر نسبت به cnn پیشنهادی
- 2: دقت بالاتر نسبت به knn پیشنهادی
- 3: برخلاف cnn در ماکزیمم های محلی گیر نمی کند.
- 4: با انتخاب یک جدا کننده مرز خوب به دقت های بالایی دست پیدا میکند
- 5: نسبت به سه روش knn معمولی و cnn معمولی (ANN) و svm معمولی دارای دقت بالاتری است .
- 6: برخلاف الگوریتم knn ما علاوه بر طبقه بندی منطقه مورد نظر غده سرطانی را نیز می توانیم مشخص کنیم

معایب [20,21]:

- 1: پیچیدگی زیاد نسبت به روش knn
- 2: زمان زیادی را برای یادگیری صرف می کند
- 3: حافظه زیادی استفاده می کند
- 4: محدودیت های ذاتی دارند مثلاً برای انتخاب پارامتر ها هنوز راه حل دقیقی وجود ندارد
- 5: به شدت به مرز جدا کننده خود و الگوریتم استخراج خود وابسته است

: بهبود شبکه عصبی :

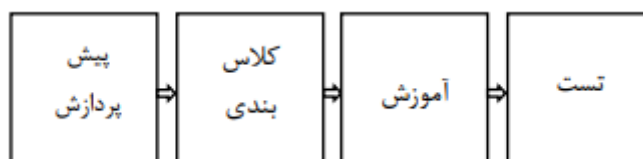
شبکه های عصبی که برگرفته از سیستم شبکه عصبی بدن انسان است از عناصر به هم پیوسته ایی به نام نورون تشکیل شده است در این مقاله از معماری خاصی از شبکه های عصبی به نام شبکه های عصبی کانولوشن اسفاده خواهیم کرد شبکه های عصبی کانولوشن نسبت به شبکه های عصبی ساده دارای مزیت خارج کردن ویژگی به صورت خودکارند. تفاوت شبکه عصبی کانولوشن با شبکه عصبی در این است که شبکه عصبی تصاویر دوبعدی را و توسط نورون ها به لایه مخفی داده میشود هر نورن به تمام نورون های قبلی خود متصل است. در حالی که شبکه کانولوشن تصاویر سه بعدی را دریافت می کند (به تصاویر سه بعدی تنسور هم گفته می شود) نورون های این شبکه به تمام لایه های ماقبل خود متصل نیستند و این مسئله باعث سرعت در اجرای الگوریتم میشود



تفاوت شبکه عصبی با شبکه عصبی کانولوشن

شبکه های عصبی کانولوشن شامل لایه های تماماً متصل (fc)¹⁵ و پولینگ¹⁶ است. [10 , 19 , 20]

در اینجا از دو لایه کانولوشن¹⁷ و یک لایه تماماً متصل استفاده میکنیم
 مراحل روش پیشنهادی در این مقاله شامل 1: پیش پردازش 2 : کلاس بندی 3: آموزش 4: تست می باشد که بلوک دیاگرام آن در شکل (6) نمایش داده شده است.



شکل 6

در این معماری تصاویری با ابعاد ثابت در ورودی مورد نیاز می باشد، به همین دلیل در پیش از آموزش تصاویر به ابعاد ثابت 257×257 تغییر اندازه داده میشود و با تغییر فرمت تصاویر سرطان سینه از PGM به JPG تصاویر برای کلاس بندی آماده میشود

¹⁵ Fully connected layer

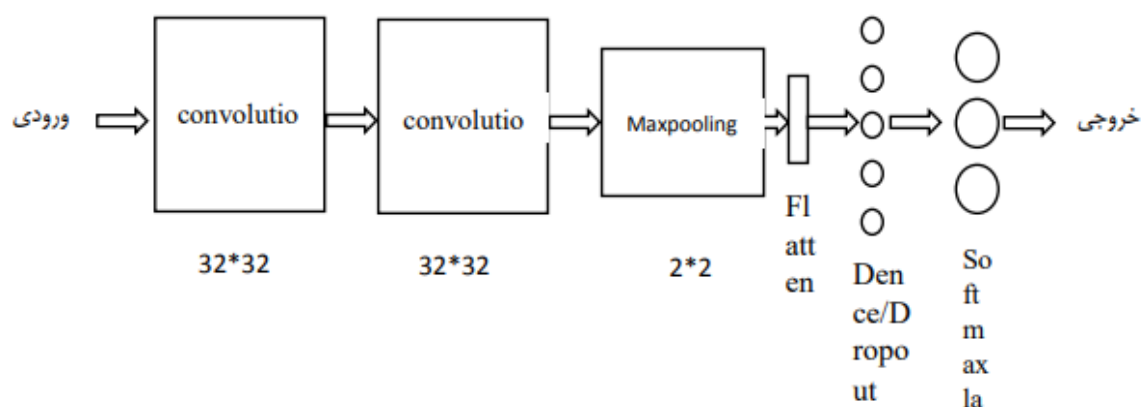
¹⁶ pooling layer

¹⁷Convolutional layer

در مرحله کلاس بندی سه کلاس برای تصاویر سرطان سینه (خوشخیم، بدخیم و نرمال) در نظر گرفته می شود.

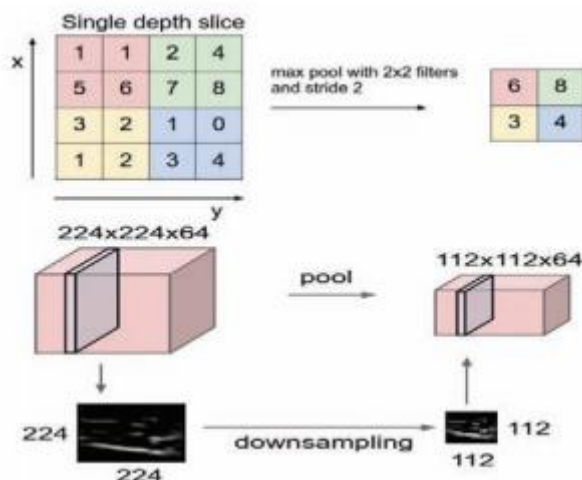
در مرحله آموزش، تصاویر آموزشی به همراه کلاس خروجی آن برای شبکه در نظر گرفته میشود. حین آموزش وزنه های فیلتر در هر تکرار به روز میشوند. به نحویکه بعد از چندین تکرار وزنه های بهینه به روز میشوند و شبکه آموزش میبندد تا بهترین ویژگی را از تصاویر استخراج کند. پس از آموزش وزنه های بهینه برای مرحله طبقه بندی و تست مورد استفاده قرار میگیرد

در این جا ما از دو لایه کانولوشن استفاده کرده ایم در درون لایه های کانولوشن نرون وجود دارد که عملیات کانولوشن را روی ورودی ها انجام می دهد مهمترین پارامتر در نرون های کانولوشن اندازه فیلتر میباشد. بعد از لایه کانولوشن از لایه ادغام استفاده میشود لایه ادغام برای کاهش ابعاد و اندازه تصویر (فقط عرض و ارتفاع نه عمق) و همچنین سرعت بخشیدن به الگوریتم مفید است و منجر به کاهش محاسبات می شود . مدل پیشنهادی بر پایه شبکه عصبی عمیق در شکل (7) نمایش داده شده است..



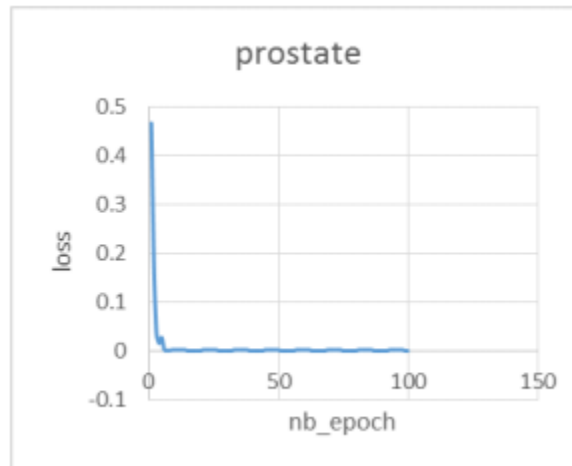
مدل پیشنهادی برای شبکه عصبی

در این مقاله برای استفاده از لایه ادغام از یکی از متداولترین لایه های ادغام به نام Max pooling استفاده شده است در شبکه های عصبی بعد از هر لایه ی کانولوشن یک لایه پولینگ داریم [1,19] که یک فیلتر با ابعاد 2×2 در نظر گرفته شده است که بیشتر عملیات در آن انجام میشود. (شکل 7)



عملکرد لایه پولینگ با فیلتر 2×2

و با در نظر گرفتن یک لایه Flatten خروجی لایه های کانولوشن که یک تانسور چند بعدی است به یک تانسور یک بعدی تبدیل میشود. شبکه عصبی عمیق از چند لایه برای درک بخشهایی از داده ها استفاده می-کند، اما برای این که بفهمیم تابع ما چقدر خوب عمل می کند نیاز به یک معیار به نام loss (مقدار خطا) نیاز داریم در این پژوهش با استفاده از softmax به محاسبه loss تصاویر آموزشی و تست پرداخته شده است.



میزان loss

که مقادیر احتمالات را در یک محدوده استاندارد نرمالیزه میکند. از تکنیک dropout به منظور جلوگیری از overfitting استفاده میشود. و در نهایت با استفاده از adam optimizer که برای محاسبه بسیاری از توابع بهینه ساز در است بهینه سازی وزن ها را انجام می دهد .

مزایا :

- 1: بالاترین دقت را نسبت همه روش ها دارد .
- 2: نسبت به سه روش knn معمولی و cnn معمولی (ANN) و svm معمولی دارای دقت بالاتری است
- 3: همانند svm علاوه بر طبقه بندی می توانیم مکان شکل مورد نظر را نیز مشخص کنیم

معایب :

- 1: احتمال گیر افتادن در ماکسیم محلی
- 2: سرعت پایین تر نسبت به دو روش بالا و پیچیدگی بالاتر

نتیجه گیری :

در این مقاله سه الگوریتم برای تشخیص سرطان سینه معرفی شد و برای بهبود این الگوریتم ها راه هایی پیشنهاد شد. که برای به دست آوردن ویژگی های مهم از hough استفاده کردیم و برای هر کدام راه هایی به صورت مختص آن روش برای بهبود آن دادیم هر کدام دارای دقت های خوب و بالایی بودند اما دیدیم از بین این سه الگوریتم ، الگوریتم cnn دارای دقت بالاتری نسبت به سایر الگوریتم ها است دقت ها را با هم بررسی میکنیم به صورت خلاصه :

الگوریتم	درصد
knn	80.4
ماشین بردار پشتیبانی (svm)	85.9
شبکه عصبی (cnn)	85

الگوریتم پیشنهادی	درصد
knn	91.1
ماشین بردار پشتیبانی (svm)	94
شبکه عصبی (cnn)	95.81

با این حال همچنان این الگوریتم ها دارای خطاهایی هستند که ما می توانیم با تکنیک هایی مثل افزایش داده¹⁸ دقت هایی بهتر و منحصر بفرد تری را به دست بیاوریم که نیاز به زمان بیشتر و همچنین سخت افزار های قدرتمند تری است

¹⁸ data augmentation

منابع :

- 1: saber fooladi ,Hassan Farsi , Sajad Mohamadzadeh Detection of Malignancy Degree in Prostate and Breast Cancers by Using Deep Neural Network 2018
- 2: R. Vijayarajeswari ,P. Parthasarathy,S. Vivekanandan ,A. Alavudeen Basha Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform d 23 May 2019
- 3: Rabat, Morocco Optimization of K-NN algorithm by clustering and reliability coefficients: application to breast-cancer diagnosis 2018
- 4: Oskouei, R. J., Kor, N. M., & Maleki, S. A. (2017). Data mining and medical world: breast cancers' diagnosis, treatment, prognosis and challenges. American journal of cancer research, 7(3), 610.
- 5: Y. A. L. Alsabahi , Lei Fan and Xiaoyi Feng Image Classification Method in DR Image Based on Transfer Learning 2018 IEEE
- 6: Spanhol FA, Oliveira LS, Petitjean C, Heutte L. A dataset for breast cancer histopathological image classification. IEEE Transactions on Biomedical Engineering. 2016 Jul;63(7):1455-62.
- 7: DeSantis, C. E., Ma, J., Goding Sauer, A., Newman, L. A., & Jemal, A. (2017). Breast cancer statistics, 2017, racial disparity in mortality by state. CA: a cancer journal for clinicians, 67(6), 439-448.
- 8: Duncan, J. S., & Ayache, N. (2000). Medical image analysis: Progress over two decades and the challenges ahead. IEEE transactions on pattern analysis and machine intelligence, 22(1), 85-106

9: Mueller, M. L. (2012). Data Mining Methods for Medical Diagnosis. Technical University of Munich.

10: Abbass, H. A. (2002). An evolutionary artificial neural networks approach for breast cancer diagnosis. Artificial intelligence in Medicine, 25(3), 265-281.

11: Diz, J., Marreiros, G., & Freitas, A. (2016). Applying Data Mining Techniques to Improve Breast Cancer Diagnosis. Journal of medical systems, 40(9), 203.

12: Chaurasia, V., & Pal, S. (2017). Performance Analysis of Data Mining Algorithms for Diagnosis and Prediction of Heart and Breast Cancer Disease.

13: Printz C. Breast cancer detection rate with ultrasound comparable to mammography. Cancer. 2016 May 15;122(10):1475.

14: "Cancer facts and figure", <http://www.cancer.org/Cancer/BreastCancer/DetailedGuide/breast-cancerkey-statistic> (2011)-(2011)

15: <http://www.mammoimage.org/databases/>

16: P. Parthasarathy, S. Vivekanandan, Investigation on uric acid biosensor model for enzyme layer thickness for the application of arthritis disease diagnosis, Health Inf. Sci. Syst. 6 (2018) 1–6.

17 : Bhatia, N. (2010). Survey of nearest neighbor techniques. arXiv preprint arXiv:1007.0085.

18: Bailey, T., & Jain, A. K. (1978). A note on distance-weighted k -nearest neighbor rules. IEEE Transactions on Systems, Man, and Cybernetics, (4), 311-313.

19 Ali Karsaz /Assistant Professor, Electrical and Bioelectric Engineering Department, Khorasan Institute of Higher Education, Mashhad Roshan/ M.Sc. Electrical and Bioelectric Engineering Department, Khorasan Institute of Higher Education, Medical image processing using deep convolutional neural networks 2020

20 : wikipedia.org

21 www.bigdata.ir

22 بهبود دقت الگوریتم KNN در داده کاوی با استفاده از قوانین وابستگی ، مهدی مرادیان ،