





عنوان پروژه 2:

تشخیص سرطان دهان به وسیله یادگیری ماشین

نام استاد راهنما:

دکتر الهام محمودزاده

نام دانشجو:

حسین رضائی

تشکر و قدردانی

مَنْتَ خدای را عز و جل که طاعتش موجب قرب است و به شکر اندرش مزید نعمت، هر نفسی که فرو می رود

ممدّ حیات است و چون بر می آید مفرّح ذات. خدای را شاگردم که این فرصت را داشتم تا در طی دوران کارشناسی

بتوانم مهارت های لازم را کسب کنم تا شاید با پیشرفت خود بتوانم ذره ای از زحمات خانواده ام را جبران کرده باشم. از

خانواده عزیزم که همواره پشتیبان و حامی بنده در تمامی مراحل زندگی ام بوده اند تشکر و قدردانی می کنم. همچنین، از

خانم دکتر محمودزاده که در طی دوره کارشناسی همواره راهنما و الگو بنده بوده اند تشکر ویژه دارم، امیدوارم بتوانم به

نحو احسن زحمات ایشان را جبران کنم.

از تمامی دوستانی که نامشان در اینجا ذکر نشد ولی همواره پشتیبان بنده بوده اند و آقایان آرتین مجد، شکیب یزدانی و

محمد دربان باران که در این پایان نامه کمک هایشان همراه بنده بود، بسیار سپاسگزارم.

فهرست مطالب

فهرست تصاویر	شش
چکیده	1
فصل اول: مقدمه	2
فصل دوم: مفاهیم پایه	4
2.1. کتابخانه ها	4
2.1.1. کراس	4
2.1.2. تنسورفلو	4
2.1.2.1. کراس و تنسورفلو 2.0	5
2.1.3. NumPy	6
2.1.4. Seaborn	7
2.2. ابزارها	7
2.2.1. گوگل درایو (Google Drive)	7

8	2.2.2 گوگل کولب (Colaboratory)
9	2.3 مدل ها
9	2.3.1 VGG16
10	2.3.2 InceptionV3
11	2.3.3 ResNet50
13	2.3.4 Xception
14	2.3.5 NasNet
18	2.3.6 Inception-Resnet-V2
19	2.3.7 Available Models
20	2.4 Optimizer ها
20	2.4.1 SGD
21	2.4.2 RMSprop
22	2.4.3 Adam
22	2.4.4 AdaMax
24	2.5 Receiver Operating Characteristic (ROC)

24 Specificity و Accuracy ,Precision ,Recall تفاوت	2.5.1
24 Specificity	2.5.1.1
24 Sensitivity یا بازیابی	2.5.1.2
24 Accuracy یا صحت	2.5.1.3
25 Precision یا دقت	2.5.1.4
25 ناحیه های مثبت و منفی	2.5.2
26 فرمول ها	2.5.3
27 فصل سوم: توضیح پروژه	
27 Image Acquisition	3.1
27 Database Generation	3.2
27 Data Augmentation	3.3
28 Classification	3.4
29 Result	3.5
31 Related Work	3.6
35 منابع	

فهرست تصاویر

9	تصویر 1: ساختار VGG16
10	تصویر 2: Different VGG Configuration
11	تصویر 3: A high-level diagram of the InceptionV3
13	تصویر 4: معماری ResNet50 و نمودار جریان مدل یادگیری عمیق (a, b و c)
14	تصویر 5: Overall Architecture of Xception (Entry Flow > Middle Flow > Exit Flow)
15	تصویر 6: Scalable Architectures for CIFAR-10 and ImageNet
	تصویر 7: Controller model architecture for recursively constructing one block of a
16	convolutional cell
16	تصویر 8: Schematic diagram of the NASNet search space
17	تصویر 9: NASNet
19	تصویر 10: ساختار Inception-Resnet-V2
20	تصویر 11: نوسان SGD

چکیده

سرطان سلول سنگفرشی دهان و دندان (OSCC) یک نوع متداول از سرطان اپیتلیوم دهان است. علیرغم تأثیر زیاد آنها بر مرگ و میر، روشهای غربالگری مناسب برای تشخیص زودهنگام OSCC اغلب فاقد دقت هستند. بنابراین OSCC ها اکثراً در مراحل آخر تشخیص داده می شوند. تشخیص زودهنگام و برآورد طرح کلی دقیق OSCC منجر به نتیجه بهتر درمانی و کاهش نرخ عود پس از درمان جراحی می شود. اغلب کارهایی که در این زمینه به کمک هوش مصنوعی انجام شده است به وسیله یادگیری ماشین می باشد. اما ما در این پایان نامه، به وسیله یادگیری عمیق و الگوریتم ها و فریم ورک های مربوط به آن به پیاده سازی یک برنامه می پردازیم که به کمک آن روی یک دیتاست که دارای تصاویر هیستوپاتولوژیکی دهان است Binary Classification انجام می دهیم (طبقه بندی ما شامل دو کلاس به نام های کلاس نرمال و کلاس سرطانی می باشد). و به تشخیص سرطان دهان می پردازیم. ما توانستیم به کمک مدل Xception و AdaMax Optimizer به $Accuracy = 90$ و $Recall = 90$ برسیم.

فصل اول

مقدمه

سرطان دهان به عنوان یک چیز زائد یا زخم در دهان ظاهر می شود که از بین نمی رود. سرطان دهان، که شامل

سرطان های لب، زبان، گونه ها، کف دهان، کام نرم و سخت، سینوس ها و حلق (گلو) می شود، در صورت عدم

تشخیص و درمان زودرس می تواند تهدید کننده ی زندگی باشد.

به طور کلی، سرطان دهان به دسته ای از بیماری های سرطانی مربوط به ناحیه ی سر و گردن گفته می شود که منحصراً

دو ناحیه ی فضای دهانی و اوروفارنکس را مورد هدف قرار می دهند.

یادگیری ماشین، مطالعه ی علمی الگوریتم ها و مدل های آماری مورد استفاده ی سیستم های کامپیوتری است که به جای

استفاده از دستورالعمل های واضح، از الگوها و استنباط برای انجام وظایف سود می برند.

به عنوان زیر مجموعه ای از هوش مصنوعی، الگوریتم های یادگیری ماشین یک مدل ریاضی بر اساس داده های نمونه یا

داده های آموزش، به منظور پیش بینی یا تصمیم گیری بدون برنامه ریزی آشکار، ایجاد می کنند.

ما در پروژه ی 1، مفهوم، تاریخچه و کاربرد های فعلی هوش مصنوعی در دندان پزشکی و کارها و تحقیقاتی که در این

زمینه انجام گرفته است را مورد بررسی قرار دادیم. در پروژه 2، هدف ما کامل کردن پروژه 1 و ادامه ی آن مفهوم می

باشد. ما در این پایان نامه به وسیله ی یادگیری ماشین، شبکه های عصبی و پردازش تصویر، به تشخیص سرطان دهان

(Oral Cancer) می پردازیم.

فصل دوم

مفاهیم پایه

2.1. کتابخانه ها

2.1.1. کراس

Keras یک API یادگیری عمیق است که در Python نوشته شده و در top پلت فرم یادگیری ماشین

TensorFlow اجرا می شود. با تمرکز بر امکان آزمایش سریع توسعه داده شده است.

2.1.2. تنسورفلو

TensorFlow یک کتابخانه نرم افزار متن باز و رایگان برای برنامه نویسی dataflow و differentiable

در طیف وسیعی از وظایف است. یک کتابخانه ریاضی نمادین است ، و همچنین برای برنامه های یادگیری ماشین مانند

شبکه های عصبی استفاده می شود و هم برای تحقیق و هم برای تولید در Google استفاده می شود.

TensorFlow توسط تیم Google Brain برای استفاده داخلی در Google ساخته شده است. با مجوز Apache

2.0 2.0 در تاریخ 9 نوامبر 2015 منتشر شد.

2.1.2.1. کراس و تنسورفلو 2.0

TensorFlow 2.0 یک پلتفرم یادگیری ماشین متن باز end-to-end است. شما می توانید آن را به عنوان

یک لایه زیرساخت برای برنامه نویسی differentiable تصور کنید که این چهار توانایی کلیدی را با هم ترکیب می

کند:

1. انجام کارآمد عملیات تنسور سطح پایین بر روی CPU، GPU یا TPU.
2. محاسبه شیب arbitrary differentiable expressions.
3. Scaling computation برای بسیاری از دستگاه ها (به عنوان مثال ابر رایانه Summit در آزمایشگاه ملی Oak Ridge، که شامل 27000 GPU است).
4. صادر کردن برنامه ها ("نمودارها") به external runtime ها مانند سرورها، مرورگرها، تلفن همراه و دستگاه های امبدد.

Keras یک API سطح بالا TensorFlow 2.0 است: یک رابط قابل قبول و بسیار پربازده برای حل مسائل یادگیری

ماشین، با تمرکز بر یادگیری عمیق جدید که abstraction های اساسی و building block ها برای توسعه و ارسال

راه حل های یادگیری ماشین با سرعت تکرار بالا را فراهم می کند.

Keras مهندسان و محققان را قادر می سازد تا از قابلیت مقیاس پذیری و کراس پلتفرم TensorFlow 2.0 به طور

کامل استفاده کنند: شما می توانید Keras را با TPU یا cluster های بزرگ GPU اجرا کنید و می توانید مدل های

Keras خود را برای اجرا در مرورگر یا تلفن همراه صادر کنید.

NumPy.2.1.3

یک بسته نرم افزاری قابل افزودن به پایتون است که کاربرد اصلی اش در مقاصد علمی و برای کار با اعداد است.

پایتون به صورت پیش فرض تنها از آرایه ها و متغیرها برای عملیات ریاضی ساده پشتیبانی می کند. بسته نام پای ویژه کار با

اعداد از راه ماتریس ها و آرایه های چندبعدی طراحی شده است. از ویژگی های آرایه ها در نام پای این است که می توان

اندازه آن ها به صورت پویا تغییر داد که این امر به افزایش سرعت برنامه نویسی کمک می کند. نام پای را می توان بسته

بنیادی پایتون برای محاسبات علمی دانست، این بسته افزون بر فراهم آوردن قابلیت کار با آرایه های ان-بعدی،

عملگرهای درایه به درایه و عملگرهای اصلی جبر خطی، قابلیت لفاف پیچی کدهای سی، سی ++ و فورترن را ممکن

می سازد.

با استفاده از آرایه‌های ان‌بعدی (ndarray) نام‌پای، می‌توان بر محدودیت‌های لیست‌های پایتون (list) که تنها با استفاده از حلقه‌های تکرار می‌توان بر روی آن‌ها کار کرد، غلبه نمود و بازدهی را بالا برد. تنها محدودیت مهم آرایه‌های ان‌بعدی نام‌پای در مقایسه با لیست‌های پایتون در این است که باید حتماً نوع داده‌های موجود در درایه‌های آن یکسان باشند. در مقابل سرعت انجام عملیاتی که با استفاده از آرایه‌های ان‌بعدی اجرا می‌شود بیشتر است.

Seaborn .2.1.4

یک کتابخانه بسیار مفید مصورسازی داده در پایتون محسوب می‌گردد. کتابخانه Seaborn روی Matplotlib ساخته می‌شود و قابلیت‌های پیشرفته بسیاری در رابطه با مصورسازی داده ارائه می‌کند.

2.2. ابزارها

2.2.1. گوگل درایو (Google Drive):

یک سرویس میزبانی فایل است که توسط گوگل ارائه می‌شود. در این سرویس کاربر می‌تواند اطلاعات خود

را در ابر قرار دهد، آنها را به اشتراک بگذارد، و اسناد خود را نیز ویرایش کند. این سرویس از ۲۴ آوریل ۲۰۱۲ (۵)

اردیبهشت ۱۳۹۱) در دسترس قرار گرفت. این سرویس شامل افزوده‌های گوگل داکس، گوگل شیتس و گوگل

اسلایدز به همراه سری برنامه‌های اداری است که به کاربر توانایی ویرایش انواع فایل‌های سرور (های ابری) گوگل را، از

جمله اسناد کتبی، صفحه گسترده‌ها، اسلایدها، نقاشی‌ها و غیره می‌دهد. از جمله ویژگی‌های گوگل درایو سینک

(هماهنگ) بودن آن با سرویس‌های دیگر از جمله: گوگل پلاس، گوگل مپس، کلاس روم، جیمیل، پیکاسا، گوگل

داکز، گوگل فوتو، بلاگر و دیگر سرویس‌های گوگل است.

همچنین این سرویس به هر کاربر ۱۵ گیگابایت حافظه اولیه می‌دهد و کاربر در صورت نیاز بیشتر می‌تواند با پرداخت

مبلغی به صورت ماهیانه، ظرفیت‌های بالاتری را گوگل درایو داشته باشد. در ماه اکتبر ۲۰۱۴، ۲۴۰ میلیون کاربر

فعال داشت.

2.2.2. گوگل کولب (Colaboratory):

یک محیط نوت‌بوک همراه با سخت‌افزار رایگان است که تماماً در فضای ابری اجرا می‌شود. کولب توسط

شرکت گوگل با هدف فراهم‌سازی شرایط کار و پیشرفت در هوش مصنوعی برای همه راه‌اندازی شد. سرویس ابری

کولب به شما امکان استفاده از پردازنده‌های GPU TPU قدرتمند برای اجرای الگوریتم‌های هوش مصنوعی بر پایه

پایتون را می‌دهد. با سرویس ابری کولب، شما و همکاران‌تان می‌توانید بدون نیاز به سخت‌افزار قدرتمندی، روی یک

پروژه کار کنید. کار با سرویس ابری کولب شبیه به Google Docs هست و همه کارها به صورت اینترنتی انجام

می‌شود. بسیاری از کتابخانه‌های محبوب هوش مصنوعی مانند TensorFlow Keras PyTorch OpenCV

Numpy Scikit-learn به صورت پیش‌فرض در کولب نصب هست و بی‌دغدغه می‌توانید از آنها استفاده کنید.

2.3. مدل ها

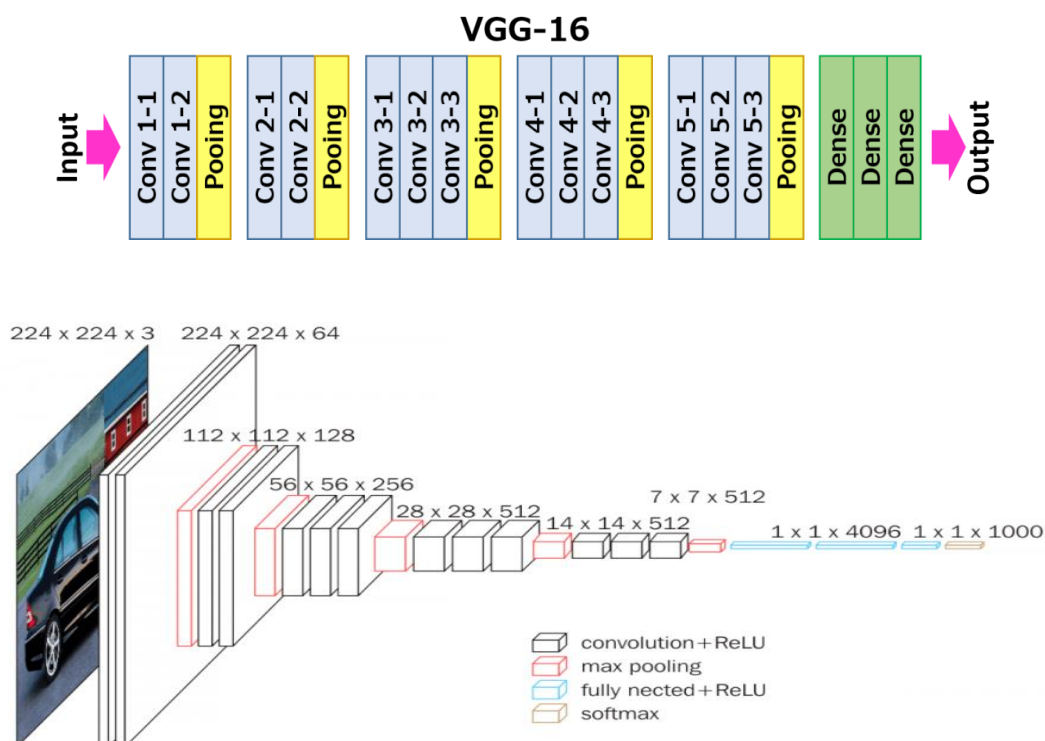
2.3.1. VGG16

توسط «Visual Graphics Group | VGG» در دانشگاه آکسفورد توسعه داده شده است و در مقاله‌ای

با عنوان «شبکه‌های عصبی خیلی عمیق برای بازشناسی تصویر بزرگ مقیاس (Very Deep Convolutional

Networks for Large-Scale Image Recognition) که در سال ۲۰۱۴ منتشر شده، توضیح داده شده است. به

طور پیش فرض، مدل انتظار تصاویر ورودی رنگی را دارد که به اندازه 224×224 بازمقیاس دهی شوند.



تصویر 1: ساختار VGG16

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224×224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

تصویر 2: Different VGG Configuration

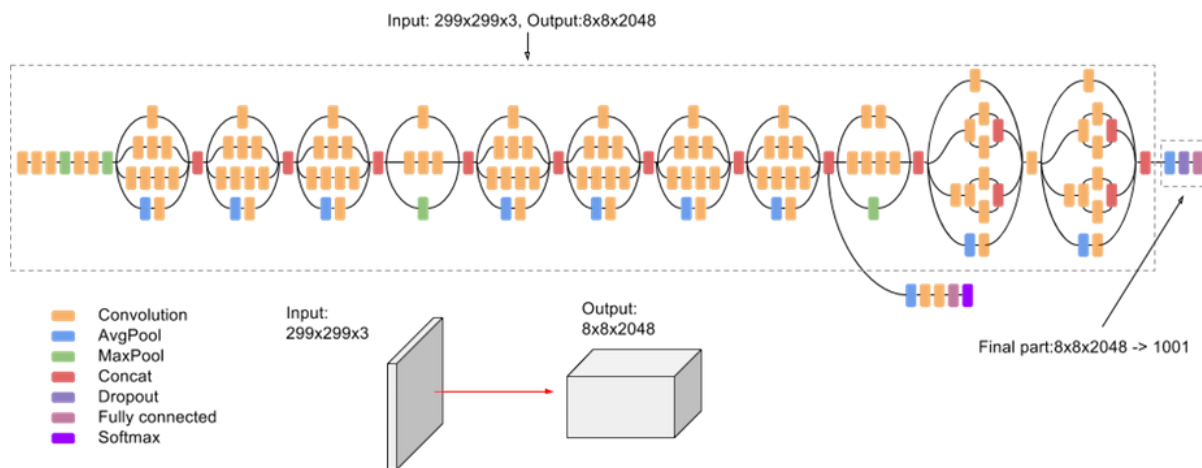
InceptionV3 .2.3.2

سومین تکرار از معماری Inception است که در ابتدا برای مدل GoogLeNet توسعه پیدا کرده بود. این

مدل توسط توسعه‌دهندگان در گوگل توسعه پیدا کرده و در مقاله‌ای با عنوان «بازبینی معماری ادراکی برای بینایی

کامپیوتری (Rethinking the Inception Architecture for Computer Vision)» توصیف شده است.

این مدل انتظار دارد که تصاویر به صورت مربعی و در اندازه 299×299 باشند. مدل را می توان به صورت زیر بارگذاری کرد.



تصویر 3: A high-level diagram of the InceptionV3

2.3.3 ResNet50

شبکه باقیمانده یا به طور خلاصه ResNet، یک مدل است که از مازول باقیمانده شامل اتصالات میان‌بر،

استفاده می کند. این مدل توسط توسعه دهندگان مایکروسافت توسعه داده شده و در مقاله‌ای با عنوان «یادگیری باقیمانده

عمیق برای بازشناسی تصویر» در سال ۲۰۱۵ منتشر شده است. مدل نیاز به تصاویر رنگی با ابعاد 224×224 دارد.

در تصویر 4، a و b: معماری ResNet50 نشان داده شده است و شامل لایه های کانولوشن، لایه های max pooling

و یک لایه fully connected است. C: یک مدل ResNet50 بر روی یک میلیون تصویر از پایگاه داده

ImageNet آموزش دیده است و می تواند تصاویر را در 1000 دسته شی طبقه بندی کند. بر اساس این مجموعه داده

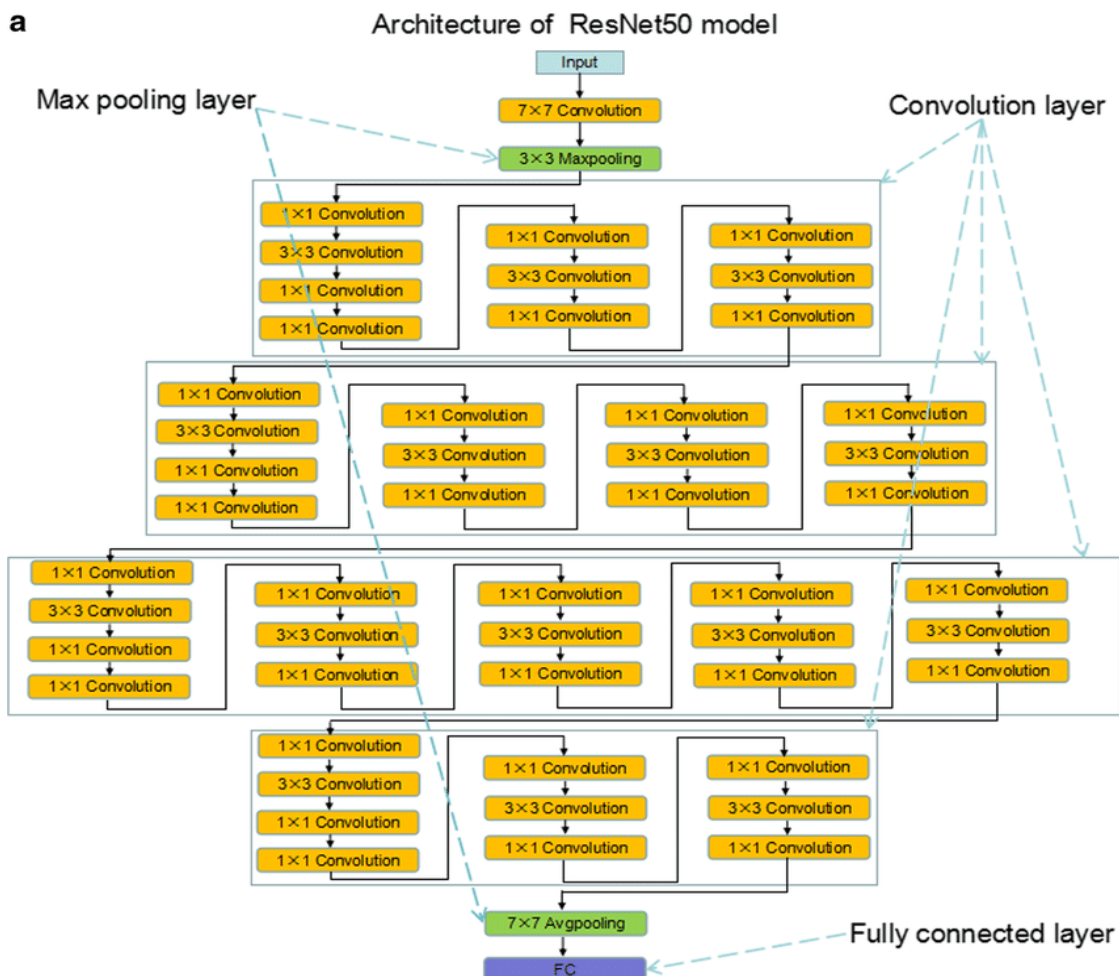
جدید از تصاویر CT، یک مدل یادگیری انتقال برای کوتاه شدن زمان آموزش و بهبود دقت مورد استفاده قرار گرفت.

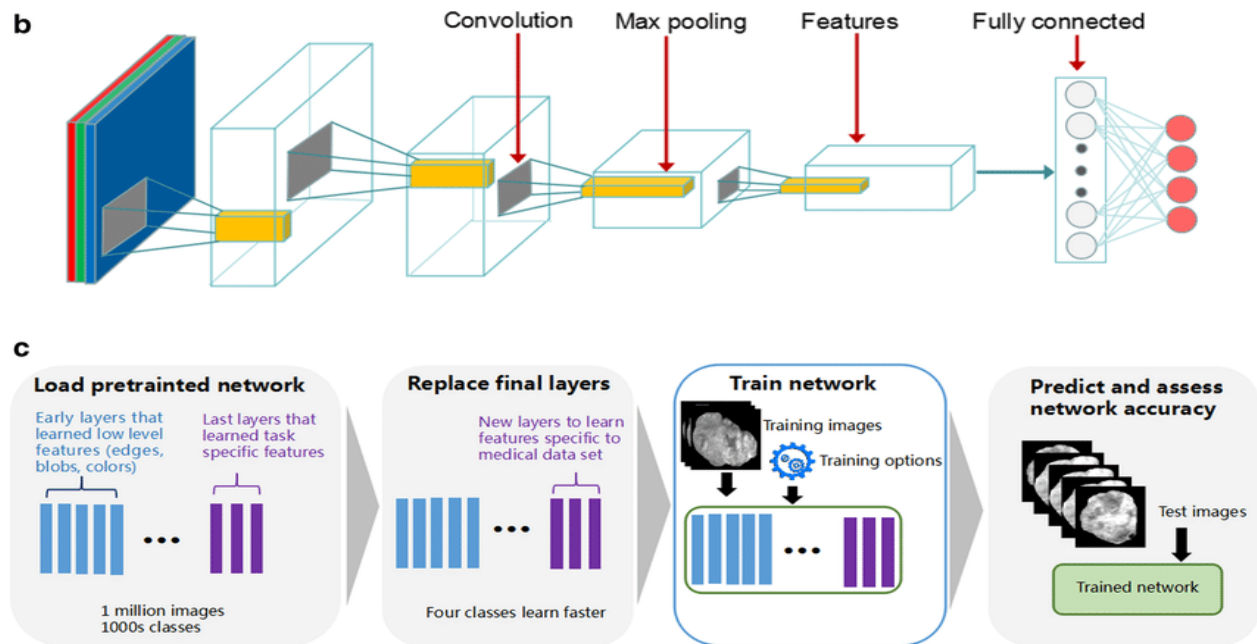
لایه های متصل قبلی بسته شدند (1 تا 174)، و لایه های متصل نهایی جایگزین شدند (175 تا 177). سرانجام، این

مدل به یک شبکه جدید تبدیل شد. پس از مشخص شدن Patch های ROI از تصاویر CT، تمام Patch ها در روش

پیشنهادی افزایش یافت. برای پیش بینی پاسخ به درمان TACE از یک شبکه عصبی کانولوشن 50 residual لایه

یادگیری انتقال استفاده شد.





تصویر 4: معماری ResNet50 و نمودار جریان مدل یادگیری عمیق. (a، b و c)

Xception 2.3.4

مدل Xception توسط Francois Chollet ارائه شده است. Xception یک توسعه از معماری

Inception است که جایگزین ماژول های استاندارد Inception با convolution های قابل تفکیک عمقی می شود.

معماری Xception: داده ها ابتدا از طریق جریان ورودی، سپس از طریق جریان میانی که هشت بار تکرار می شود و در

نهایت از طریق جریان خروجی عبور می کنند. توجه داشته باشید که تمام لایه های Convolution و

SeparableConvolution با نرمال سازی دسته ای دنبال می شوند (در نمودار موجود نیست). تمام لایه های

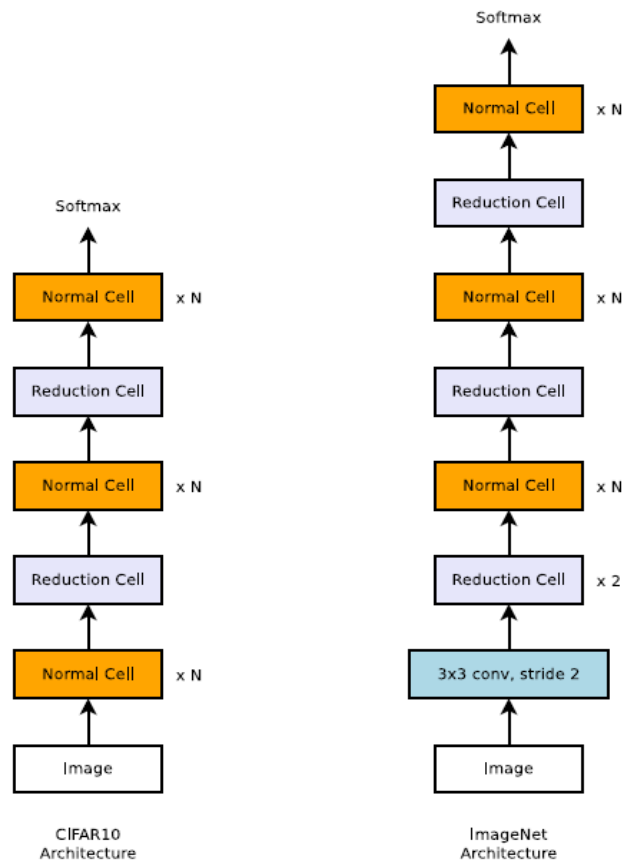
SeparableConvolution از ضرب عمق 1 استفاده می کنند (بدون گسترش عمق).

NasNet.2.3.5

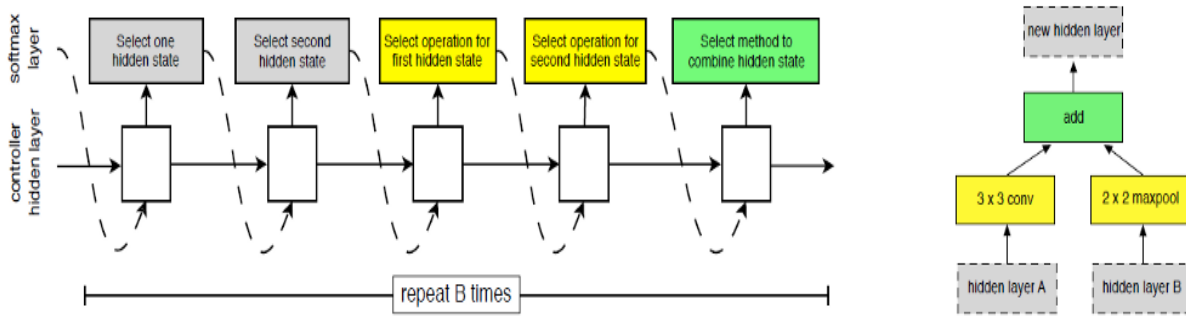
سلول ، این سلول را به ImageNet اعمال می کنند. یک تکنیک منظم سازی جدید به نام ScheduledDropPath

نیز ارائه شده است که به طور قابل توجهی تعمیم در مدل های NASNet را بهبود می بخشد. سرانجام ، مدل NASNet

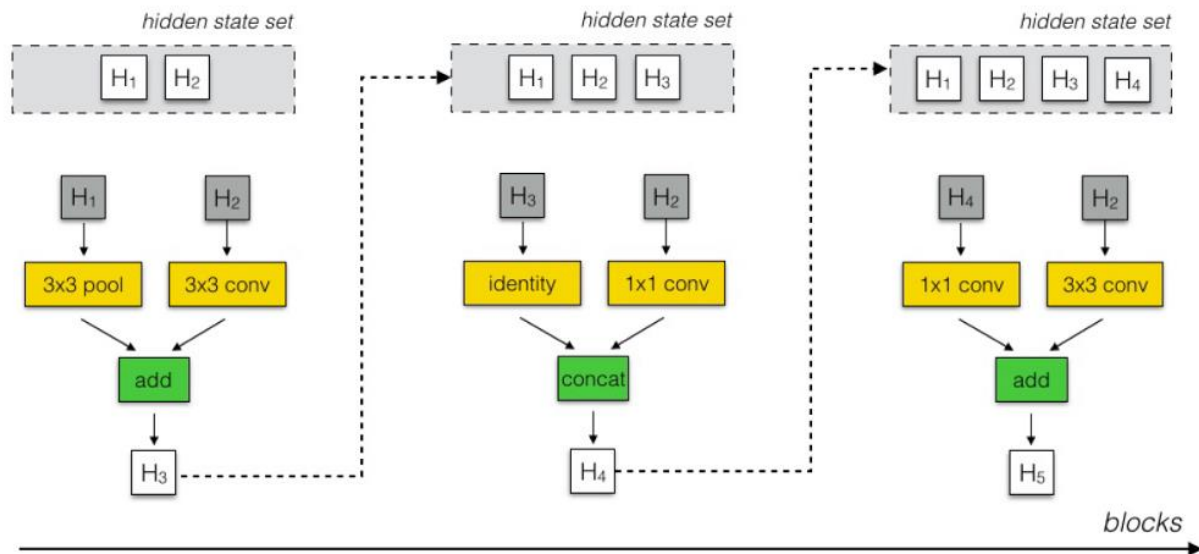
با اندازه مدل کوچکتر و پیچیدگی کمتر (FLOP) به نتایج پیشرفته دست می یابد.



تصویر 6: Scalable Architectures for CIFAR-10 and ImageNet

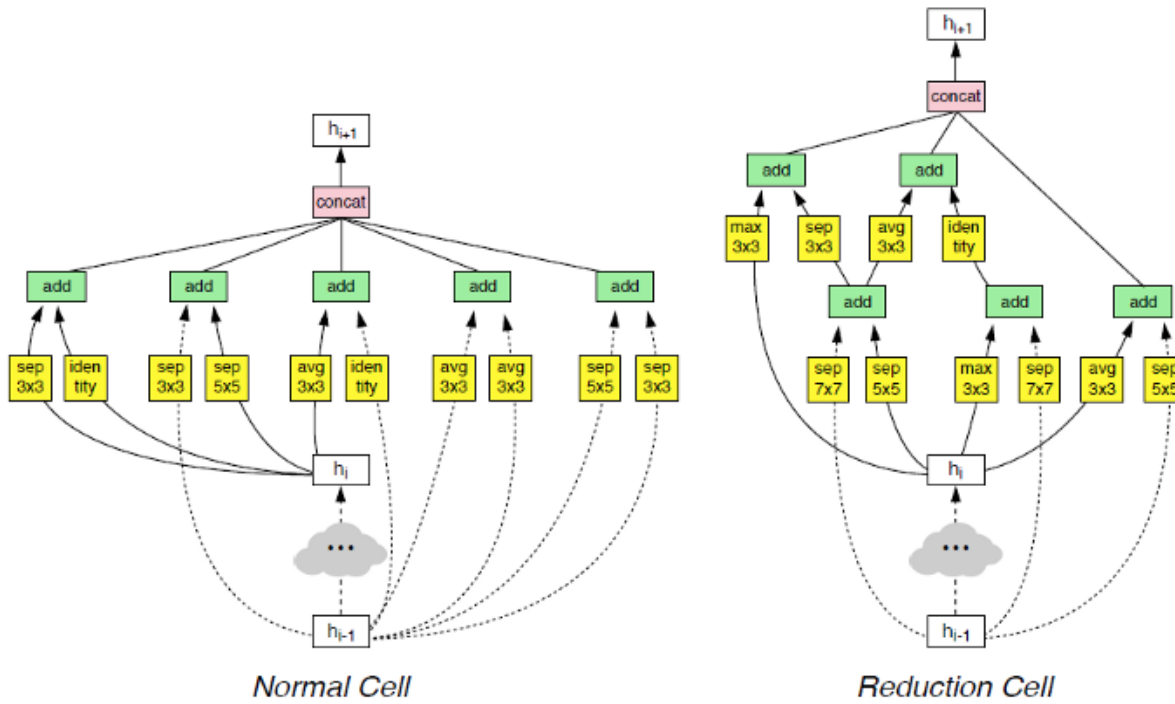


تصویر 7: Controller model architecture for recursively constructing one block of a convolutional cell

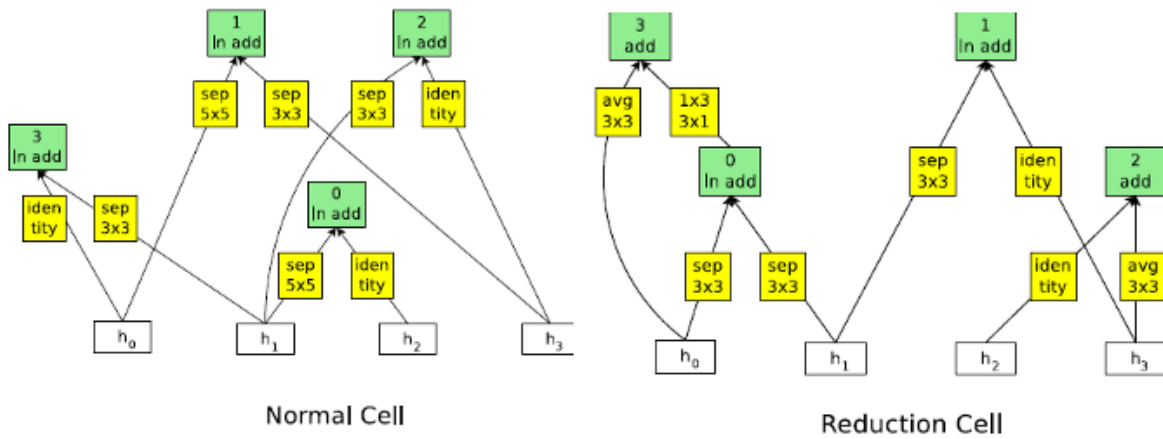


تصویر 8: Schematic diagram of the NASNet search space

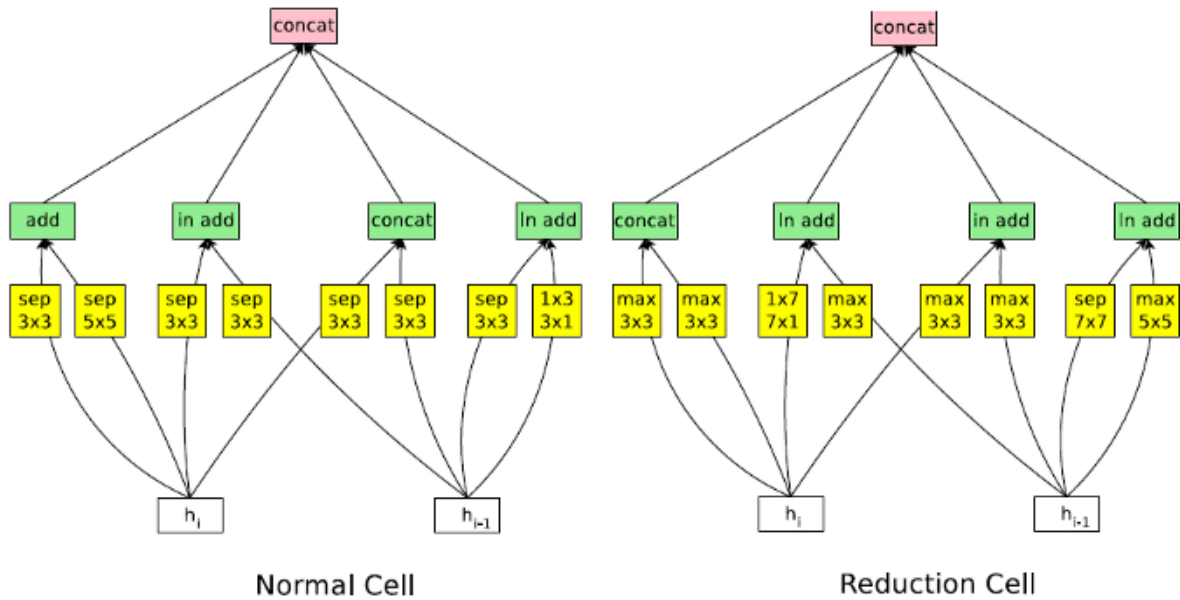
NASNet-A



NASNet-B (4 inputs & 4 outputs)



NASNet-C

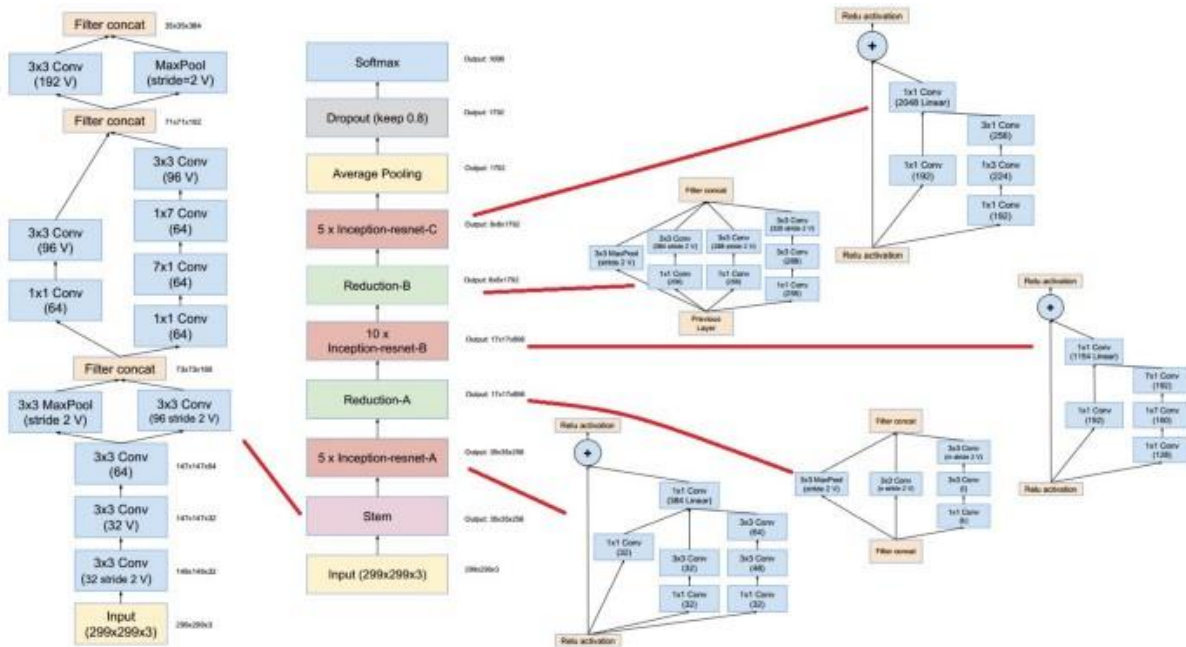


تصویر 9: NASNet

Inception-Resnet-V2 .2.3.6

هر دو شبکه Inception و Residual از معماری SOTA هستند که عملکرد بسیار خوبی با هزینه محاسباتی

نسبتاً کم نشان داده اند. Inception-ResNet این دو معماری را برای تقویت بیشتر عملکرد ترکیب می کند.



تصویر 10: ساختار Inception-Resnet-V2

2.3.7. Available models

Model	Size	Top-1 Accuracy	Top-5 Accuracy	Parameters	Depth
Xception	88 MB	0.790	0.945	22,910,480	126
VGG16	528 MB	0.713	0.901	138,357,544	23
VGG19	549 MB	0.713	0.900	143,667,240	23
ResNet50	98 MB	0.749	0.921	25,636,712	-
InceptionV3	92 MB	0.779	0.937	23,851,784	159
InceptionResNetV2	215 MB	0.803	0.953	55,873,736	572
NASNetLarge	343 MB	0.825	0.960	88,949,818	-

*The top-1 and top-5 accuracy refers to the model's performance on the ImageNet validation dataset.

**Depth refers to the topological depth of the network. This includes activation layers, batch normalization layers etc.

2.4. Optimizer ها

2.4.1. SGD

Stochastic gradient descent یک بروزرسانی پارامتر را برای هر نمونه آموزش $x(i)$ و برچسب $y(i)$

انجام می دهد:

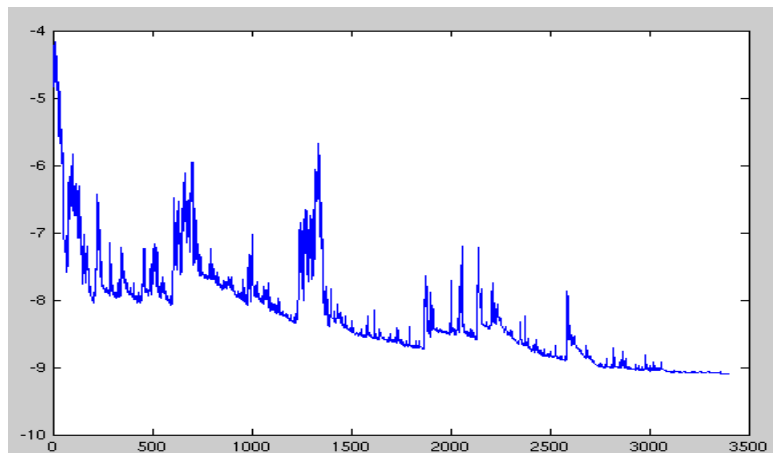
$$\theta = \theta - \eta \cdot \nabla_{\theta} J(\theta; x^{(i)}; y^{(i)})$$

نزول شیب دسته ای، محاسبات اضافه را برای مجموعه داده های بزرگ انجام می دهد، زیرا شیب نمونه های مشابه را قبل

از هر بروزرسانی پارامتر محاسبه می کند. SGD با انجام همزمان یک به روزرسانی، این افزونگی را از بین می برد.

بنابراین معمولاً بسیار سریعتر است و می توان از آن برای یادگیری آنلاین نیز استفاده کرد. SGD به روزرسانی های مکرر

را با یک واریانس زیاد انجام می دهد که باعث می شود عملکرد هدف به شدت در تصویر 11 نوسان داشته باشد.



تصویر 11: نوسان SGD

در حالی که نزولی شیب دسته ای به حداقل basin ای که پارامترها در آن قرار دارند همگرا می شود ، نوسان SGD از یک طرف آن را قادر می سازد تا به حداقل های محلی جدید و بالقوه بهتر پرش کند که از طرف دیگر ، نهایتاً همگرایی را به حداقل ممکن پیچیده می کند ، زیرا SGD، overshooting را ادامه می دهد. با این حال ، نشان داده شده است که وقتی سرعت یادگیری را به آرامی کاهش می دهیم ، SGD رفتار همگرایی یکسان با نزول شیب دسته ای نشان می دهد ، تقریباً به ترتیب به حداقل محلی یا سراسری برای بهینه سازی غیر محدب و محدب تبدیل می شود.

RMSprop 2.4.2

RMSprop یک روش نرخ یادگیری انطباقی منتشر نشده است که توسط جف هینتون ارائه شده است.

RMSprop و Adadelata هر دو به طور مستقل و در یک زمان ساخته شده اند که ناشی از نیاز به حل و فصل کاهش

میزان یادگیری Adagrad است.

$$E[g^2]_t = 0.9E[g^2]_{t-1} + 0.1g_t^2$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{E[g^2]_t} + \varepsilon} g_t$$

RMSprop همچنین نرخ یادگیری را با میانگین نمایی شیب مربع در حال تجزیه تقسیم می کند. هینتون پیشنهاد می کند

γ روی 0.9 تنظیم شود ، در حالی که مقدار پیش فرض خوبی برای میزان یادگیری η ، 0.001 است.

Adam.2.4.3

Adaptive Moment Estimation (Adam) روش دیگری است که نرخ یادگیری انطباقی را برای هر

پارامتر محاسبه می کند. علاوه بر ذخیره متوسط انحراف نمایی شیب های مربع قبل v_t مانند Adadelta و RMSprop

، آدام همچنین از میانگین انحراف نمایی شیب های گذشته m_t ، مانند شتاب حفظ می کند. در حالی که حرکت می

تواند به عنوان یک توپ در حال حرکت به سمت پایین شیب دیده شود ، آدام مانند یک توپ سنگین با اصطکاک رفتار

می کند ، بنابراین حداقل های مسطح را در سطح خطا ترجیح می دهد. میانگین انحراف شیب های مربع گذشته و گذشته

m_t و v_t را به ترتیب زیر محاسبه می کنیم:

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

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

مقادیر پیشنهادی: $\epsilon = 10^{-8}$ ، $\beta_1 = 0.9$ و $\beta_2 = 0.999$ است.

AdaMax .2.4.4

فاکتور v_t در قانون به روزرسانی آدام ، گرادیان را به نسبت معکوس با نرم ℓ_2 شیب های گذشته (از طریق

v_{t-1} و شیب فعلی $|g_t|^2$ مقیاس می کند:

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

ما می‌توانیم این به روزرسانی را به نرم l_p تعمیم دهیم. Kingma و Ba، β_2 را به صورت β_2^p پارامتر می‌کنند:

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

نرم‌ها برای مقادیر p بزرگ از نظر عددی ناپایدار می‌شوند، به همین دلیل نرم‌های l_1 و l_2 در عمل بیشتر رایج هستند.

با این حال، l_∞ همچنین به طور کلی رفتار پایداری از خود نشان می‌دهد. به همین دلیل، نویسندگان AdaMax را

پیشنهاد کرده‌اند (Kingma و Ba، 2015) و نشان می‌دهند که v_t با l_∞ به مقدار پایدار زیر همگرا می‌شود. برای

جلوگیری از سردرگمی با آدم، از u_t برای نشان دادن v_t با محدودیت نرم بی‌نهایت استفاده می‌کنیم:

$$u_t = \beta_2^\infty v_{t-1} + (1 - \beta_2^\infty) |g_t|^\infty = \max(\beta_2 \cdot v_{t-1}, |g_t|)$$

اکنون می‌توانیم با جایگزینی $\varepsilon + \sqrt{\hat{v}_t}$ با u_t این مورد را در معادله بروزرسانی آدم قرار دهیم تا قانون بروزرسانی AdaMax را بدست آوریم:

$$\theta_{t+1} = \theta_t - \frac{\eta}{u_t} \hat{m}_t$$

توجه داشته باشید که همانطور که u_t به عملکرد MAX متکی است، بایاس نسبت به صفر به اندازه m_t و v_t در Adam

قابل پیشنهاد نیست، به همین دلیل نیازی به محاسبه تصحیح بایاس برای u_t نداریم. مقادیر پیش فرض خوب

$\beta_1 = 0.9$ ، $\eta = 0.002$ و $\beta_2 = 0.999$ است.

2.5. Receiver operating characteristic (ROC)

2.5.1. تفاوت Specificity و Accuracy, Precision, Recall

2.5.1.1. Specificity

عبارت است از کسری از جواب‌های منفی که به درستی تشخیص داده شده‌است مثلاً درصد افرادی که طبق

پیش‌بینی مدل سرطان ندارند و در دنیای واقعی هم سرطان ندارند.

2.5.1.2. Recall یا Sensitivity یا بازیابی

عبارتست از کسری از جوابهای مثبت که درست تشخیص داده شده‌اند مثلاً درصد افرادی که طبق پیش‌بینی

مدل سرطان دارند و در دنیای واقعی هم سرطان دارند.

2.5.1.3. Accuracy یا صحت

یعنی مقدار اندازه‌گیری شده چقدر به مقدار واقعی نزدیک است برای accuracy باید precision بالا باشد

ولی برعکسش لزوماً برقرار نیست. بالا بودن بایاس و واریانس به معنای accuracy کم است.

2.5.1.4 Precision یا دقت

برای اندازه گیری های متوالی از یک مقدار میزان نزدیک بودن مقادارهای اندازه گیری را نشان می دهد. مثلاً اگر

یک ساعت هر روز فقط ۲ ساعت جلو رود مقدار accuracy یا صحتش پایین است ولی مقدار precision آن

بالاست.

2.5.2. ناحیه های مثبت و منفی

واقعاً درست یا **condition positive (P)**: تعداد موارد واقعاً مثبت در داده

واقعاً منفی یا **condition negative (N)**: تعداد موارد واقعاً منفی در داده

مثبت درست یا **true positive (TP)**: معادل موارد درست

منفی درست یا **true negative (TN)**: معادل مردود کردن درست

مثبت نادرست یا **false positive (FP)**: معادل با خطای نوع اول (مثبت کاذب- فرض صفر درست باشد و آزمون فرض

آن را رد کند.) و خطای نوع دوم (منفی کاذب- فرض صفر درست نباشد و آزمون فرض آن را قبول کند.)

منفی نادرست یا **false negative (FN)**: معادل با خطای نوع اول و دوم

2.5.3. فرمول ها

sensitivity, recall, hit rate, or true positive rate (TPR)

$$\text{TPR} = \frac{\text{TP}}{\text{P}} = \frac{\text{TP}}{\text{TP} + \text{FN}} = 1 - \text{FNR}$$

precision or positive predictive value (PPV)

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} = 1 - \text{FDR}$$

accuracy (ACC)

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{P} + \text{N}} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

F1 score (is the harmonic mean of precision and sensitivity)

$$\text{F}_1 = 2 \cdot \frac{\text{PPV} \cdot \text{TPR}}{\text{PPV} + \text{TPR}} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}$$

فصل سوم

توضیح پروژه

Image Acquisition .3.1

Data Set این پروژه از مقاله ی Rahman T.Y., Lipi B. Mahanta, Histopathological imaging “ (2020). “ database for oral cancer analysis گرفته شده است داده ها از 230 بیمار به دست آمده است.

طبقه اول داده ها دارای رزولوشن 100x و طبقه بعدی داده ها دارای رزولوشن 400x هستند هر دو طبقه دارای دو کلاس

به اسم های تصاویر نرمال و تصاویر سرطانی هستند. ما با تصاویر با رزولوشن 400x کار می کنیم که کلاس اول مربوط

به تصاویر نرمال به تعداد 201 عکس و کلاس دوم مربوط به تصاویر سرطانی به تعداد 495 عکس است.

Database Generation .3.2

داده ها را به صورت تصادفی به 80٪ و 20٪ تقسیم می کنیم. 80٪ برای داده های آموزش (395 عکس سرطانی و 162

عکس نرمال) و 20٪ برای تست (100 عکس سرطانی و 39 عکس نرمال).

Data Augmentation .3.3

برای افزایش تعداد داده ها (Data Augmentation) روی عکس های مربوط به آموزش یک بار Flip-right-left و یک بار Flip-top-bottom انجام می دهیم لذا در نهایت تعداد تصاویر نرمال آموزش برابر 486 و تعداد تصاویر سرطانی آموزش برابر 1185 می باشد.

Classification 3.4

ما در این پایان نامه از مدل های مختلفی به اسم های VGG16، VGG19، ResNet50، NASNetLarge، InceptionV3، Xception و InceptionResNetV2 و همچنین از Optimizer های مختلف به اسم های Adam، AdaMax، RMSprop و SGD استفاده کرده ایم که در ادامه نتایج مربوط به هر کدام را مقایسه می کنیم. ابتدا مدل InceptionV3 را با $\text{Epoch} = 10$ روی Optimizer ها با Learning Rate های مختلف امتحان می کنیم. با توجه به جدول نتایج که در ادامه آمده است، نمودار های حاصل، مقادیر پیش نهاد شده Learning rate توسط ارائه دهندگان Optimizer ها و توضیحاتی که برای هر Optimizer در بخش قبل داده شد به این نتیجه می رسیم که AdaMax Optimizer با $\text{Learning Rate} = 0.001$ برای پروژه ما مناسب تر می باشد.

حال مجدداً برای اطمینان بیشتر همین فرآیند را با $\text{Epoch} = 100$ انجام می دهیم طبق جدول نتایج که در ادامه آمده است تقریباً به همان نسبت نتایج قبل می رسیم. حال AdaMax Optimizer را با $\text{Learning Rate} = 0.001$ و $\text{Epoch} = 100$ برای مدل های مختلف انجام می دهیم. که نتایج زیر حاصل می شود:

Result .3.5

Input_Shape	Model	Optimizer	Epochs	Accuracy	Recall_0 Recall_1	Precision_0 Precision_1	F1-score_0 F1-score_1
(384,512,3)	InceptionV3	Nadam(lr=1e-3)	10	0.42	0.21 0.97	0.95 0.32	0.34 0.49
(384,512,3)	InceptionV3	Nadam(lr=0.01)	10	0.81	0.85 0.72	0.89 0.65	0.87 0.68
(384,512,3)	InceptionV3	Nadam(lr=0.001)	10	0.88	0.92 0.79	0.92 0.79	0.92 0.79
(384,512,3)	InceptionV3	Nadam(lr=0.0001)	10	0.88	0.90 0.85	0.94 0.77	0.92 0.80
(384,512,3)	InceptionV3	SGD(lr=0.001)	10	0.72	1.00 0.00	0.72 0.00	0.84 0.00
(384,512,3)	InceptionV3	SGD(lr=1e-3)	10	0.79	0.97 0.33	0.79 0.81	0.87 0.47
(384,512,3)	InceptionV3	RMSprop(lr=1e-5)	10	0.76	0.76 0.74	0.88 0.55	0.82 0.63
(384,512,3)	InceptionV3	RMSprop(lr=1e-4)	10	0.88	0.94 0.72	0.90 0.82	0.92 0.77
(384,512,3)	InceptionV3	RMSprop(lr=1e-3)	10	0.9	0.92 0.85	0.94 0.80	0.93 0.83
(384,512,3)	InceptionV3	RMSprop(lr=1e-2)	10	0.89	0.92 0.82	0.93 0.80	0.92 0.81
(384,512,3)	InceptionV3	RMSprop(lr=0.001)	10	0.84	0.83 0.87	0.94 0.67	0.88 0.76
(384,512,3)	InceptionV3	Adamax(lr=1e-3)	10	0.9	0.93 0.82	0.93 0.82	0.93 0.82
(384,512,3)	InceptionV3	Adamax(lr=0.001)	10	0.83	0.91 0.62	0.86 0.73	0.88 0.67
(384,512,3)	InceptionV3	Adam(lr=1e-3)	10	0.28	0.00 1.00	0.00 0.28	0.00 0.44
(384,512,3)	InceptionV3	Adam(lr=0.001)	10	0.72	1.00 0.00	0.72 0.00	0.84 0.00
(384,512,3)	InceptionV3	Adamax(lr=1e-3)	100	0.9	0.92 0.85	0.94 0.80	0.93 0.83
(384,512,3)	InceptionV3	Adamax(lr=0.001)	100	0.9	0.92 0.85	0.94 0.80	0.93 0.83
(384,512,3)	InceptionV3	Adamax(lr=0.001)	100	0.88	0.89 0.85	0.94 0.75	0.91 0.80
(384,512,3)	InceptionV3	Adamax(lr=0.001)	150	0.91	0.92 0.90	0.96 0.81	0.94 0.85
(384,512,3)	InceptionV3	Adamax(lr=0.001)	250	0.9	0.91 0.87	0.95 0.79	0.93 0.83

(384,512,3)	InceptionV3	Adamax(lr=0.001)	350	0.89	0.89 0.90	0.96 0.76	0.92 0.82
(384,512,3)	InceptionV3	RMSprop(lr=0.001)	100	0.89	0.92 0.82	0.93 0.80	0.92 0.81
(384,512,3)	InceptionV3	Adam(lr=10 ⁻⁸)	100	0.87	0.87 0.87	0.95 0.72	0.91 0.79
(384,512,3)	InceptionV3	SGD(lr=1e-3)	100	0.83	0.93 0.56	0.85 0.76	0.89 0.65
(192,256,3)	InceptionResNetV2	Adamax(lr=0.001)	100	0.86	0.89 0.79	0.92 0.74	0.90 0.77
(192,256,3)	NASNetLarge	Adamax(lr=0.001)	100	0.72	1.00 0.00	0.72 0.00	0.84 0.00
(192,256,3)	VGG16	Adamax(lr=0.001)	100	0.76	0.86 0.51	0.82 0.59	0.84 0.55
(192,256,3)	VGG19	Adamax(lr=0.001)	100	0.78	0.86 0.56	0.83 0.61	0.85 0.59
(384,512,3)	ResNet50	Adamax(lr=0.001)	100	0.9	0.91 0.87	0.95 0.79	0.93 0.83
(384,512,3)	Xception	RMSprop(lr=1e-3)	100	0.86	0.92 0.69	0.88 0.77	0.90 0.73
(384,512,3)	Xception	Adamax(lr=1e-3)	100	0.91	0.93 0.87	0.95 0.83	0.94 0.85
(384,512,3)	Xception	Adamax(lr=0.001)	100	0.94	0.94 0.92	0.97 0.86	0.95 0.89
(384,512,3)	Xception	Adamax(lr=0.001)	100	0.91	0.92 0.90	0.96 0.81	0.94 0.85
(384,512,3)	Xception	Adamax(lr=0.001)	250	0.9	0.90 0.90	0.96 0.78	0.93 0.83
(192,256,3)	Xception	Adamax(lr=0.001)	250	0.91	0.92 0.87	0.95 0.81	0.93 0.84
(192,256,3)	Xception	Adamax(lr=0.001)	500	0.88	0.90 0.85	0.94 0.77	0.92 0.80

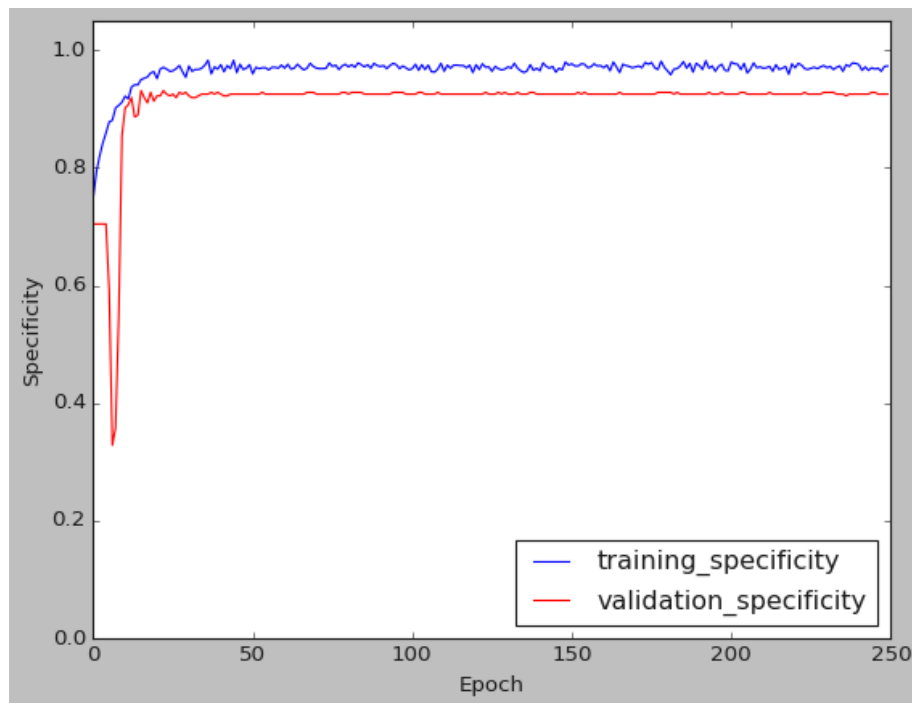
با توجه به بخش Available Models در فصل قبل برای هر مدل و نتایج جدول بالا، دو مدل InceptionV3 و

Xception دارای نتایج بهتری نسبت به مدل های دیگر هستند. لذا این دو مدل را با Epoch بیشتر نسبت به حالت قبل

(250) اجرا می کنیم. همان طور که می بینید ردیف 3 تا مانده به آخر (Model , Input_Shape = (384,512,3),

(= Xception, Optimizer = AdaMax (lr = 0.001), Epoch = 250, Accuracy = 90, Recall = 90)

در نهایت دارای بهترین نتیجه است که نمودار آن به شکل زیر می باشد.



در بعضی از موارد به دلیل سنگین بودن شبکه و عدم توان در اجرای آن، برای Input Shape مجبور به استفاده از

ورودی (192,256,3) به جای (384,512,3) شده ایم. اما برای دیدن تاثیر اندازه ی Input Shape بر روی نتایج،

مدل Xception با AdaMax Optimizer را یک بار با ورودی (192,256,3) و یک بار با ورودی

(384,256,3) اجرا کرده ایم با توجه به جدول نتایج همان طور که می بینید ورودی (384,256,3) دارای نتایج بهتری

نسبت به ورودی (192,256,3) است.

3.6. Related work

Reference	Learning approach	Study sample	Methods	Experimental outcome
Chodorowski et al. (Chodorowski et al., 2000b)	Machine learning	True color images of the mouth or oral cavity.	Automated grading of oral lesions, oral lichenoid reactions and oral leukoplakia using Fisher's linear discriminant, kNN-Nearest, Gaussian quadratic, Neighbor and Multilayer Perceptron.	Linear discriminant showed the best classification accuracy.
Ray et al. (Ray et al., 2008)	Machine learning	Histological Oral Sub-mucosal Fibrosis images	Automated segmentation of OSF images to constituent layers: background, epithelial and connective tissue using hybrid segmentation algorithm.	The hybrid algorithm showed better accuracy than the region growing algorithm.
Krishnan et al. (M. Krishnan et al., 2010)	Machine learning	Histological Oral Sub-mucosal Fibrosis images	Automated grading of Oral Sub-mucosal Fibrosis using wavelet-texture features fed into SVM and Bayesian classifier	SVM classifier with linear kernel function provides the highest classification accuracy of 92% than the Bayesian classifier.
Krishnan et al. (M. M. R. Krishnan et al., 2011)	Machine learning	Histological Oral Sub-mucosal Fibrosis images	Automated grading of Oral Sub-mucosal Fibrosis without dysplasia (OSFWD) and Oral Sub-mucosal Fibrosis with dysplasia using Higher-Order Spectra (HOS) and Local Binary Pattern (LBP) to SVM as the classifier.	The highest value for Sensitivity of 94.07% and specificity of 93.33% using HOS features reported.
Patra et al. (Patra et al., 2012)	Machine learning	Histological Oral Sub-mucosal Fibrosis images	Automated grading of Oral Sub-mucosal Fibrosis using SVM and Bayesian classifier.	SVM showed better accuracy of 96.6% than the Bayesian classifier.

Reference	Learning approach	Study sample	Methods	Experimental outcome
Dev Kumar Das, Pabitra Mitra et al. 2017	machine learning	histopathological images	Computational approach for mitotic cell detection and its application in oral squamous cell carcinoma	89% precision, 95% recall or sensitivity, 97.35% specificity, 96.92% accuracy, 96.45% AUC and 92% F-score measure. random forest tree classifier.
Marc Aubreville, Christian Knipfer et al. 2017	Deep learning	Confocal Laser Endomicroscopy (CLE) images	Automatic Classification of Cancerous Tissue in Laserendomicroscopy Images of the Oral Cavity using Deep Learning.	(AUC) of 0.96 and a mean accuracy of 88.3% (sensitivity 86.6%, specificity 90%)
Dev Kumar Das, Surajit Bose et al. 2018	deep CNN	oral tissue histological images	Automatic identification of clinically relevant regions from oral tissue histological images for oral squamous cell carcinoma diagnosis	12-layered ($7 \times 7 \times 3$ channel patches).texture-based feature (Gabor filter) trained random forests. 96.88% of accuracy.
Rahman et al. (T. Y. Rahman et al., 2018)	Machine learning	Histological Oral Squamous Cell Carcinoma images	Automated binary classification or grading (normal vs. malignant) of Oral Squamous Cell Carcinoma using texture-based features fed into linear SVM as classifier.	100% classification accuracy reported
Santisudha Panigrahi, Tripti Swarnkar 2019	Deep learning	Oral Cancer Histopathology images	Automated Classification of Oral Cancer Histopathology images using Convolutional Neural Network	4-layered (5X5X3) patches of CNNs with 10 fold cross validation.achieved 96.77 % accuracy

Reference	Learning approach	Study sample	Methods	Experimental outcome
Dev Kumar Das, Subhranil Koley et al. 2019	deep CNN	oral histological images	Computer aided tool for automatic detection and delineation of nucleus from oral histopathology images for OSCC screening	*Recall– 0.88 PPV-82 F-score 0.85. ** DC- 0.94, JI-0.89, Precision-0.97, Recall-0.91
*for detection of nucleus: Wavelet downsampled patch based 12 layers CNN.(active contour; Chan-Vese model) **for segmentation of nucleus: Wavelet downsampled patch based CNN with AC-NSCT.				
Pandia et al. (Jeyaraj & Samuel Nadar, 2019)	Deep learning	hyperspectral images different from histological (biopsy) images	Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm.	accuracy of 91.4% for number of 100 images and 94.5% for number of 500 images by proposed CNN algorithm
Rahman et al. (Tabassum Yesmin Rahman et al., 2020)	Machine learning	Histological Oral Squamous Cell Carcinoma images	Automated binary classification or grading (normal vs. malignant) of Oral Squamous Cell Carcinoma using color, shape and texture-based features fed into SVM, Logistic regression, decision tree and linear discriminant as a classifier.	99.4% accuracy was reported using the decision tree, 100% accuracy using SVM and linear regression.
N. Das, E. Hussain and L.B. Mahanta (2020)	deep learning	biopsy images	Automated classification of cells into multiple classes in epithelial tissue of oral squamous cell carcinoma using transfer learning and convolutional neural network. classify OSCC into its four classes as per the Broder's system of histological grading	the proposed CNN model outperformed the transfer learning approaches displaying an accuracy of 97.5% .

- [1] Chodorowski, A., Mattsson, U., & Gustavsson, T. (2000a). Oral lesion classification using true-color images. *Doktorsavhandlingar Vid Chalmers Tekniska Hogskola*, 3661(1646), 1–12.
- [2] Chodorowski, A., Mattsson, U., & Gustavsson, T. (2000b). Oral lesion classification using true-color images. *Doktorsavhandlingar Vid Chalmers Tekniska Hogskola*, (1646), 1–12. <https://doi.org/10.1117/12.348507>
- [3] Krishnan, M., Chakraborty, C., & Ray, A. (2010). Wavelet based texture classification of oral histopathological sections. *Microscopy: Science, Technology, Applications and Education*, 897–906. Retrieved from <http://www.formatex.info/microscopy4/897-906.pdf>
- [4] Krishnan, M. M. R., Acharya, U. R., Chakraborty, C., & Ray, A. K. (2011). Automated diagnosis of oral cancer using higher order spectra features and local binary pattern: A comparative study. *Technology in Cancer Research and Treatment*, 10(5), 443–455. <https://doi.org/10.7785/tcrt.2012.500221>
- [5] Ray, M. E. T., Reddy, D. S., Mukherjee, A., Chatterjee, J., Paul, R. R., & Dutta, P. K. (2008). Detection of constituent layers of histological oral sub-mucous fibrosis: Images using the hybrid segmentation algorithm. *Oral Oncology*, 44(12), 1167–1171. <https://doi.org/10.1016/j.oraloncology.2008.03.002>

- [6] Patra, R., Chakraborty, C., & Chatterjee, J. (2012). Textural Analysis of Spinous Layer for Grading Oral Submucous Fibrosis. *International Journal of Computer Applications*, 48(22), 33–37. <https://doi.org/10.5120/7513-0563>
- [7] Rahman, T. Y., Mahanta, L. B., Chakraborty, C., Das, A. K., & Sarma, J. D. (2018). Textural pattern classification for oral squamous cell carcinoma. *Journal of Microscopy*, 269(1), 85–93. <https://doi.org/10.1111/jmi.12611>
- [8] Rahman, Tabassum Yesmin, Mahanta, L. B., Das, A. K., & Sarma, J. D. (2020). Automated oral squamous cell carcinoma identification using shape, texture and color features of whole image strips. *Tissue and Cell*, 63(December 2019), 101322. <https://doi.org/10.1016/j.tice.2019.101322>
- [9] N. Das, E. Hussain and L.B. Mahanta, Automated classification of cells into multiple classes in epithelial tissue of oral squamous cell carcinoma using transfer learning and convolutional neural network. *Neural Networks* (2020), doi: <https://doi.org/10.1016/j.neunet.2020.05.003>.
- [10] Jeyaraj, P.R., Samuel Nadar, E.R. Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm. *J Cancer Res Clin Oncol* 145, 829–837 (2019). <https://doi.org/10.1007/s00432-018-02834-7>
- [11] Das D.K., Koley S., Bose S., Computer aided tool for automatic detection and delineation of nucleus from oral histopathology images for OSCC screening, *Applied Soft Computing*, Volume 83, October 2019, 105642. <https://doi.org/10.1016/j.asoc.2019.105642>

- [12] Panigrahi S., Swarnkar T., Automated Classification of Oral Cancer Histopathology images using Convolutional Neural Network. 978-1-7281-1867-3/19/\$31.00 ©2019 IEEE.
- [13] Dasa D.K., Bose S., Automatic identification of clinically relevant regions from oral tissue histological images for oral squamous cell carcinoma diagnosis. 0040-8166/ © 2018 Published by Elsevier Ltd. <https://doi.org/10.1016/j.tice.2018.06.004>.
- [14] Aubreville M., Knipfer C., Jaremenko C., Maier A., Automatic Classification of Cancerous Tissue in Laserendomicroscopy Images of the Oral Cavity using Deep Learning. SCIENTIFIC Reports | 7: 11979 | DOI:10.1038/s41598-017-12320-8.
- [15] Das D.K., Mitra P., Chakraborty C., Computational approach for mitotic cell detection and its application in oral squamous cell carcinoma. Multidim Syst Sign Process DOI 10.1007/s11045-017-0488-6.
- [16] Rahman T.Y., Lipi B. Mahanta, Histopathological imaging database for oral cancer analysis. <https://doi.org/10.1016/j.dib.2020.105114>.