Deep Learning Based Image Classification for Remote Medical Diagnosis

Juliana Shihadeh†*, Anaam Ansari‡*, Tokunbo Ogunfunmi§*
†jshihadeh@scu.edu,‡aaansari@scu.edu,§togunfunmi@scu.edu
*Department of Electrical Engineering
Santa Clara University, Santa Clara, California 95053

***Abstract*—In this paper, we describe research on Convolutional Neural Networks to advance access to medical diagnosis. Our research focuses on classifying skin cancer images as Benign or Malignant as a starting base to build on to develop an app that allows easily-accessible medical diagnosis in underdeveloped countries. The methodology of our research is based on Convolutional Neural Networks: AlexNet [3] and GoogleNet [4].**

We hope to expand our work to be able to classify additional physically visible diseases in addition to Skin Cancer, since a camera only captures exterior physically visible diseases [5]. All our training and testing presented in this paper has been run on the NVIDIA Jetson TX2 GPU. Results are promising, showing accuracy rates up to 74 percent depending on how neural network parameters are changed. Down the road, we intend to incorporate this technology with a previously developed Vital Signs Multi- sensor kit [1]. The kit will be a compact and affordable device equipped with sensors that can be used to take a patients vital signs, such as blood pressure, heart rate, blood glucose concentration, and blood oxygen saturation. Combined, the tools will provide a complete system for remote medical diagnosis.

***Index Terms*—Machine Learning, Medical Diagnosing, Skin Cancer, Benign, Malignant, AlexNet, Convolutional Neural Networks, Deep Learning**

I. Introduction

A. Technology and Health

In many developing countries, a large percentage of the population lacks access to adequate health care. Many have very few or no doctors within miles. For example, in Africa and India close to 70 percent of the population lives in rural areas and has little to no access to hospitals or clinics [1]. Should they need to make the journey to a hospital, they often times cannot afford to pay to see a doctor. Telemedicine, a breakthrough in the past couple decades [1], has proven to break down this barrier between a patient and physician and now machine Learning [2], an increasingly growing field continuing to affect our lives in new ways, can be used to further enhance Telemedicines abilities. If a health care facility or doctor is not close by, or if it would simply save the patient a lot of time/effort/money from going to the doctor directly, machine learning based diagnosis could help bridge this gap.

Machine Learning is a growing field that continues to affect our lives and in recent years has become an even stronger tool to aid medical diagnosis. Emerging studies reveal researchers have been orchestrating new algorithms and Neural Network architectures that increase the speed of analyzing MRI scan, classifying lung diseases, and diagnosing cancers [1], [2], [3]. The goal of our research is to assist in that.

Using Convolutional Neural Network Architectures, our research presented is part of a gradually developing study toward designing an accurate yet easily accessible platform for medical diagnosing, especially to communities who lack adequate medical care. The potential merge of our Neural Network into an app, would allow patient diagnosis on the spot without needing a doctor. Our research is focused on classifying melanoma images into benign (negative) or malignant (positive) classes. We have tested and trained the AlexNet and GoogleNet on the NVIDIA Jetson TX2 to classify skin moles as cancerous or not.

B. Motivation

We would like to use the benefits of Artificial Intelligence to develop accessible medical care to any who may lack access to medical services. The ratio of doctors for every 1,000 people in 2016 in India and South Africa was 0.8 [4]. After having learned more about image recognition and understood the possibilities available with it, we decided we wanted to focus our research toward identifying medical health issues since our goal is to make a direct impact and positive change on communities that face a shortage in access to health care. Our research is to identify skin cancer images. Melanoma, if detected early on is a curable disease, that said we would like to achieve results that can identify melanoma in its early stages for anyone regardless of location and accessible resources.

The main goal of the medical diagnosis tool we are creating is to ensure it is usable in remote areas where communities are in short supply of local doctors and health facilities. We would like to turn it into an app accessible on devices available in third world countries, or sending over a few phones that the app can run on so communities can share the device among families, or via cloud based services. If the tool is accessible and usable around the world, anyone in need of

978-1-5386-5566-5/18/$31.00 ©2018 IEEE

medical diagnosis can be assisted directly on the spot. As engineers, we can harness the power of technology and science to extend services and make them low cost and accessible.

C. Health Stats

Our research is currently aimed at classifying skin cancer but the bigger goal for our research is to expand into diagnosing other diseases and illnesses. The following is a brief overview of statistics on skin cancer, showing the impact of it in underdeveloped low-income countries to signify the impact having technology that makes medical diagnosing more accessible to anyone anywhere would be. Each year 8.8 million people die from cancer, mostly in low- and middle-income countries....with less than 30 percent of low-income countries" having access to health care services for "accessible diagnosis and treatment" and in the case of cancer it is often difficult to have a professional nearby they can be referred to leading to "delayed and fragmented care that spirals into further complicated health difficulties [5]. Recently in 2015, division of accessible health care between high-income and low-income countries was further emphasized when statistics showed that "approximately 35 percent of low-income countries reported that pathology services were generally available in the public sector, compared to more than 95 percent of high-income countries [5].

As of this year only 286, 723 new cases of skin cancer" have already been "brought up across the world [6]. "The five-year survival rate for melanoma detected in its earliest states is around 97 percent," but as a result of any misdiagnosis that leads to detecting melanoma in its latest stages, the survival rate "drops to approximately 14 percent" [7]. In South Africa due to the geographical location and exposure to sunlight, skin cancer incidents are high, more especially among the white population being one of "highest incidences of malignant melanoma in the world... the concern for skin cancer overall has grown in recent years. The estimated yearly incidence of malignant melanoma is 4.76 per 100,000 persons overall and 19.2 per 100,000 in whites [8]. Making an app accessible to anyone anywhere, could assist saving thousands of lives by alerting someone of a medical issue they have before it becomes too late when a doctor couldn't be near them to do that for them.

D. Potential Technology Barriers

The big question is how would we make this technology accessible to anyone, anywhere? Since technology accessibility is limited in underdeveloped countries and we have limited knowledge on what types of accessible technology there are in each community we would work with, we are still exploring solutions that would allow our AI smart medical diagnosing software accessible in underdeveloped countries. Our goal is to upload our backend Neural Network software into an app we develop so that community members in underdeveloped countries can use the technology.

The base of the solution is to create an app. However, how to deliver the app and update the app requires more analysis. One way to go about it is to create an app that relies on a cloud, and upload our software onto the cloud. But that requires wifi which in turn requires researching the individual communities we make our app accessible so we can accommodate to their wifi access and figure out a way to connect to wifi if they don't have access to any at all. The pros of this solution is we wouldn't need to send over someone to update the software if updates are needed since it could automatically update on the cloud. The con to this solution, however, obviously is that wifi access would be a barrier we would need to devise a solution for; with the software being on a cloud, the app not only needs the wifi to make updates but also just to run. Another way to go about it, is to have all the software needed and required for the app to work without any wifi connection uploaded onto the device its on (which we discuss below about devices to send). The pros for this is the community members have an app with all the software preloaded so it can run without any wifi. The cons to this solutions is that we need to send someone over to update the software if we make updates.

In either case, we will most likely need to send devices: touch-screen phones and/or iPads. Each community would be given a few tablets, depending on population size, so that locals can use it for their medical diagnosing, with the Vital Signs Multi-Sensor Kit [1] as a medical care package. So it would almost be/become a virtual clinic. And, we would need to train a local who is willing to take on the responsibility of learning how to use the technology and acting as a "facilitator" for it to assist locals in using it.

II. Related Work

Several new studies and projects have emerged in recent years to apply Artificial Intelligence to Skin Cancer Detection.

In 2017 a team of two Software Engineers from Intel, Mike Borozdin and Peter Ma, developed an app that uses a Convolutional Neural Network that detect skin cancer with an image of a mole [9]. Ma notes the level of impact this revolutionizing technology has on society when explaining how in creating this new technology they're "basically able to do real-time screening for patients without any waiting time [9] Within 24 hours of developing the technology, Borozdin and Ma were able to boost the Network's accuracy to 85 percent.

Another study led at Stanford University was also dedicated to using Convolutional Neural Networks to detect Skin Cancer, and developing algorithms to improve the accuracy. The team of researchers developed an algorithm that outperforms Dermatologists' accuracy rates [7]. Andre Esteva and Brett Krupel, the two engineers, began experimenting with an existing Google algorithm developed to train and classify 1.28 million images of animals into 1000 classes [7]. Even though it was created to be adapted to classify between a cat and dog they applied the GoogleNet Inception Model to classify Melanoma images as benign or malignant [7]. After modifying the algorithm to alter the sensitivity control of the CNN, the CNN was able to reach an accuracy of 72.1 0.9 percent [10].

In Germany, a research team initiated a study in applying Convolutional Neural Networks to medical imaging classification that led to a collaboration with France and the U.S. Together they researched modifying CNNs to classify Skin Cancer. All three countries' work improved CNNs to achieve an accuracy of "88.9 percent in classifying malignant melanomas and 75.7 percent in classifying non-cancer moles". [11] [12].

Furthermore, AlexNet, the main model of our research, has proven to be successfully applicable to other medical application. In a recent study on classifying lung patterns and depict what, if any, lung disease are prevailing in the image [13] the research team used a CNN architecture that was a spin off of AlexNet's architecture. Furthermore, the team of two software engineers at Intel used the AlexNet for their work as well and had achieved an 85 percent accuracy. This shows us promising use of AlexNet for other medical diagnosing. With many new studies emerging to apply Artificial Intelligence to Skin Cancer Detection, it is a promising sign to support further application of AI to other medical diagnosing.

III. Background on AI

Artificial Intelligence has led to the innovation of technology smart enough to act and make decisions like a human. Under Artificial Intelligence is a subcategory identified as Machine Learning, and within Machine Learning is a subcategory identified as Deep Learning. The overarching goal of machine learning is to optimize a machines decision making skills. The three main categories are Supervised Learning, Unsupervised Learning, and Reinforcement. Our research depends on supervised learning, specifically image recognition.

A. Supervised Learning

Supervised learning involves predictions based on previous evidence from experimental procedures outside of the machine itself. It involves training with labelled data and making predictions on unlabelled data. Supervised learning includes classification and regression. Classifications is done when there exists discrete output classes and regression is done when the output classes are continuous.

A common use case of supervised learning is image recognition [14]. Image Recognition is considered to be supervised learning because it relies on giving a network labelled data and characteristics to classify unlabelled data.We will discuss Image recognition implementation in detail in section III-D

B. Unsupervised Learning

Unsupervised learning is when no previous results are present for the machine on which to make future decisions. In unsupervised learning, since the machine is not given predetermined data it must teach itself how to make decisions through observations [15]. Unsupervised learning is considered to be the closest form of learning to humans and animals since they learn through the observation of their setting [15]. Unsupervised learning achieves this observation based on a training process called Deep Learning. The machine is fed data it will need to identify, to learn how to identify this data it must train. Through the process of deep learning, the machine clusters current data it has access to in order to determine patterns within it in which the conclusions can be made based off off. Clustering methods are practice in unsupervised learning and are listed below:

- K means (choosing K is tricky, if it is too small it will be sensitive to noise points and if it is too large it will include unnecessary points
- Mixture Models
- Expectation Maximization (EM)
- Belief Propagation

C. Reinforcement

Reinforcement is the third category of machine learning, it learns from experience. The machine is given a problem which it has no correct answers to or any information as how to achieve those answers [15]. Under reinforcement, the machine then tests different methods and tracks what their results are [15]. The results it discovers to be accurate are the ones it then focuses on to backtrack and understand how it achieved that result, so that it can produce new solutions based on these methods that produce more "wins" or correct results [15]. The machine learns how to make decisions based on previous decisions it made that led it to correct results. Reinforcement includes Value-based, Policy-based, and Model-based. Common examples that use reinforcement are online games like chess and checkers. Over time, the machine learns from its own game playing and from the opponent's moves, and through experience begins to make decisions based on its previous decisions [16].

- Value-Based: Q-learning, real-time, dynamic programming
- Policy-based: Actor-Critic Algorithm
- Model-based: Games likes Chess and GO

D. Image Recognition

Image Recognition, as previously briefly explained, is a supervised learning process of a machine identifying images by training on previously labeled images. A network is fed a large set of images (typically thousands to produce accurate results) labeled under "classes" (labels of what the image is). The network then can train itself by identifying certain characteristics of the images so it knows what it needs to look for in an image to identify as a certain class. That being done, the network would be able to look at an image, correlate its characteristics with the ones its learned and find out which class is the closest match.

In the process of images recognition, a machine determines questions to identify the image such as:

1) What type of image is it?
2) What kind of objects are in the image?
3) What is the lighting, pose, context, shadow?
4) What shapes can I make through the clutter?

Training relies on Deep Learning, which relies on Neural Networks such as the Convolutional Neural Networks and

Deep Neural Neural Networks. The process of going through the layers is what is considered to be deep learning since data is being processed through multiple layers of the network to train. Convolution Neural Networks are a type of Deep Neural Network (DNN), the multiple-layers allow a network to learn the data by itself and in turn figure out how to classify data on its own. Figure 1 shows an overview of how the Convolutional Neural Network breaks down an image through multiple layers to classify it.

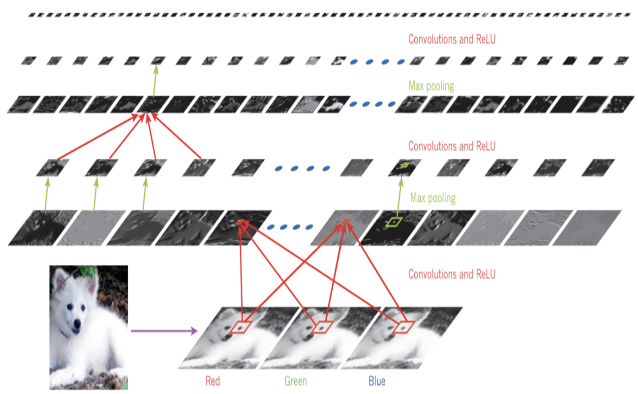


Fig. 1: Convolutional Neural Network [14]

IV. Methods and Process

A. Methods

Our research is based on supervised learning. We fine-tune two Convolutional Neural Networks: AlexNet and the GoogleNet Inception Model. The following is a brief overview of these two models.

1) AlexNet: An AlexNet is comprised of 5 Convolutional Neural Networks, 3 Fully Connected Layers and a soft max layer. In Figure 2 the AlexNet Architecture presented is divided into two so it can be processed on two GPUs for higher speed and processing capability.

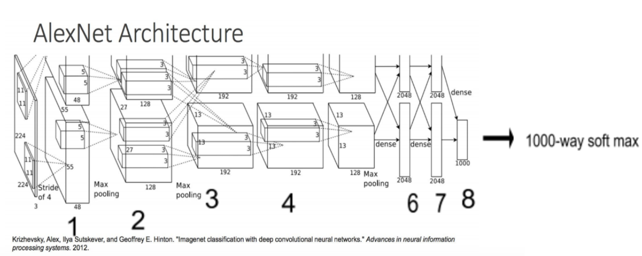


Fig. 2: The Alex Net Architecture [17]

In the AlexNet presented in Figure 2:

- Layers 1 - 5 are the Convolutional Neural Network Layers
- Layers 6-8 are the fully connected layers
- And, at the output a 1000-way softmax finalizes all results.

The final layer which is the SoftMax layer takes the values of the last fully-connected layer to generate a probability measure for each of the output classes. After obtaining the probability measure of each class, the machine then outputs the label of the class that has the highest probability which is the class that the neural networks believes to be the most likely best classification of what the image is based on its training.

For the sake of our research initially we chose to follow Kratzerts tutorial on how to create and finetune AlexNet [18] to train with an image dataset of dogs and cats that can be found on Kaggle [19] The output of the AlexNet for the cats and dogs application had two output classes (dogs and cats) as opposed to 1000, just like the output classes would be for our Skin Cancer Application. In doing so, the application was a base foundation to our research experiments so we could transition into running AlexNet on NVDIA's Jetson TX2 and training the network on medical images [20]

2) GoogleNet: The GoogleNet's architecture relies heavily on non-linearity. The more non-linearity in a neural network the higher its accuracy is. It has 22 layers, carefully designed to maximize the use of space within each layer so that images are analyzed on a more in-detail level [21]. These 22 layers, though, are only the layers that have parameters, and parameters do play one of the most vital roles in a CNNs performance [21]. However, within these 22 layers are dimensions of more layers. That said, in total there actually 100 layers in the GoogleNet [21]. Figure 3 shows a high-level image of the GoogleNet layers and Figure 4 shows a more in-depth explanation of the layers. In our research we finetune the last full layer, which in the case of the GoogleNet is Inception Layer 5 which includes all layers 5a - 5b

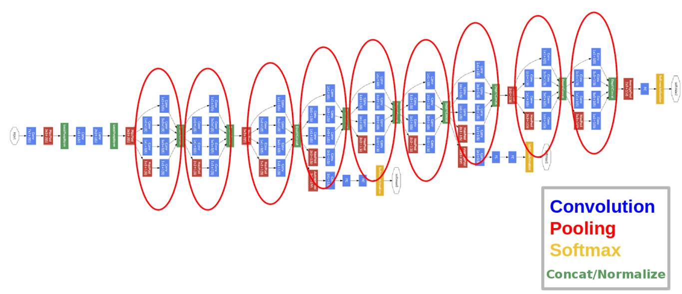


Fig. 3: The GoogleNet Architecture [22]

V. Procedure

A. Preliminary Training and Fine Tuning Using Cat and Dog Images

For the sake of initiating our research to achieve the needed steps to have a working AlexNet we could finetune, the following tools were used:

- Tensorflow
- Vi Terminal
- Python3
- Virtual Environment
- OpenCV
- Numpy
- Jupyter Notebook

978-1-5386-5566-5/18/$31.00 ©2018 IEEE

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

[21]

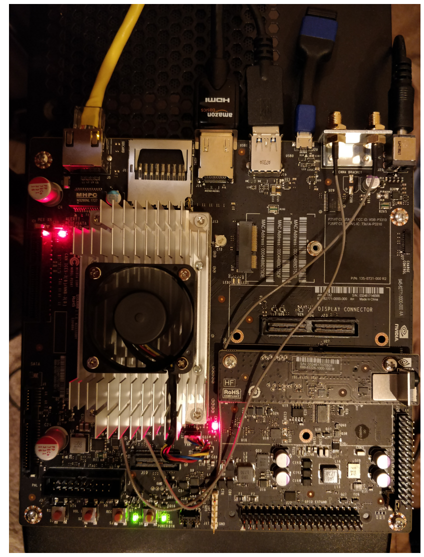
Fig. 4: A Breakdown of the GoogleNet Layers [21]

In order to finetune AlexNet, we needed to alter a segment of the CNNs to train the layer's weights on a new set of images so that AlexNet could later classify images similar to them without any human aid. As previously mentioned, we finetuned AlexNet to classify Cat and Dog images from ImageNet [23] using the ImageNet trained Caffe weights but reducing the output classes to 2 instead of 1000.

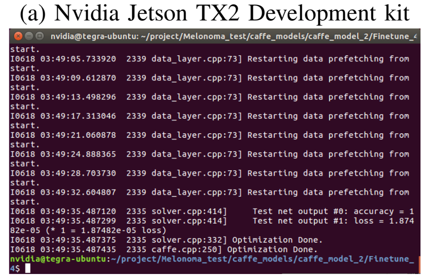
With an introduction to how AlexNet works and how it can be trained, we were able to finetune it to Skin Cancer images on the NVDIA Jetson Tx2 to train and test.

B. Training and Fine Tuning Using Medical Images

We ran AlexNet and GoogleNet on the NVDIA Jetson TX2 to combine software and hardware for the process of identifying medical images and observe how their accuracy compares to each other as well as the accuracy rates of other ongoing studies. Nvidia Jetson TX2 is Nvidia's power efficient embedded GPU deviced that is specialized for AI computing. It has 2 Nvidia Pascal architecture GPU cores and 4 ARM cores along with 8GB of RAM. In Figure 5a shows the NVIDIA Jetson TX2 development kit.



(a) Nvidia Jetson TX2 Development kit



(b) Training on Nvidia Jetson TX2

Fig. 5: NVIDIA Jetson TX2

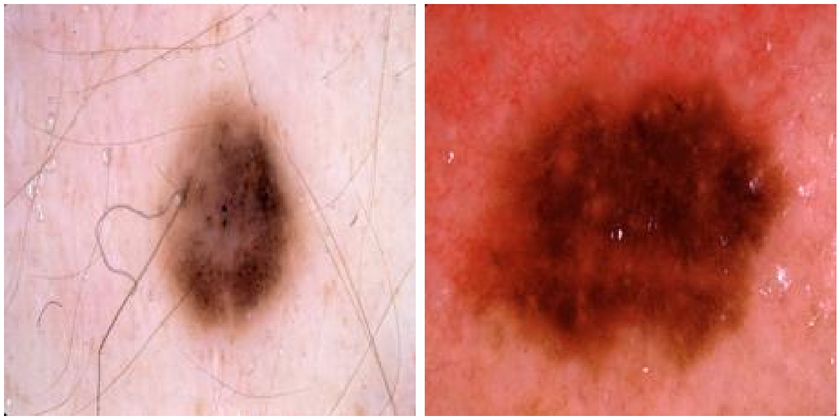
1) The Dataset:: For our research we use the International Skin Imaging Collaboration image database [24]. Cases like the research we are conducting allows us to use the database to engineer solutions, such as Convolutional Neural Networks, that can use the data to diagnose skin cancer (and more medical issues when large databases are collected for them) independent of doctors, which is especially important where there's no access to medical care. The images 6a,**??**,**??** are examples of benign moles and the images **??**,6b,**??** are examples malignant moles.

In total there were 23906 images available to work with, however not all of the images had metadata to identify if the image was a benign or malignant mole so we ran scripts to extract and save ”benign” or ”malignant” labeled images only which cut down our database to 13,025 images to work with. Most of the images were in JPG form, a few were in png form which we converted into JPG for us to work in. All of the data that classifies the images as ”benign” or ”malignant” was extracted from the metadata in the JSON files affiliated with each image.

For our research, the ISIC was the only database we were able to get access to that holds Skin Cancer Images. That said, the ISIC did only include images of cancers on light skin only. However, we dont believe that poses as much as a problem or weakness to the accuracy as is thought because the part of the images that the network focuses on to determine whether it is a cancerous mole or not is the mole itself. The mole is what holds the characteristics the neural network looks at in the image to classify it as Benign or Malignant, and the mole's features are more independent of the skin color.

We first ran scripts on all the ISIC images to weed out all images that weren't labeled as ”benign” or ”malignant”, resize the images, and convert ant non-jpg images to jpg images.

Our work relies on transfer learning, which in other words is finetuning, so the neural network trains and learns using pretrained weights, Caffe weights, on AlexNet and GoogleNet rather than training from completely from scratch [25] [26]. Before we trained the AlexNet or the GoogleNet, part of the fine-tuning process required a changing the names of the fully connect layers we were training so that new weights could be created based on those layers. Training a neural network would require a much larger database and take many more hours.



(a) Benign sample (b) Malignant sample

Fig. 6: Example of benign and malignant image scans

Using these pretrained weights as a starting point allows us to selectively choose the AlexNet layers to train on and observe the improvement of accuracy in.

After we were done running the scripts on our database to prepare the data to fit in the neural network and have the right data for the neural network to be able to learn, as well as making the necessary changes to the AlexNet and GoogleNet we would upload the skin cancer images on the NVDIA Jetson TX2 to train and test. Part of the training of the neural network requires running validation tests, and in all our runs we've reached a maximum 100 percent. Finally, in the last stage of our experiments, after the neural network finishes training we feed it images of moles it hasn't seen to classify as benign or malignant, the resulting percent the neural network prints out is the accuracy rate which is what we list in the data results. We first went through all our test trials on AlexNet before shifting to GoogleNet.

VI. Data and Results

Explanation of the variables in solver prototext [27] Of our experiments, the highest accuracy rate reached 74.59 percent. This was in the case of having 11,107 images. That said, the more images available for the Neural Network to train the higher the accuracy rate of the neural network.

1) AlexNet Test Results: Table I presents a focused analysis on a set of 844 training images and 244 images only and observe how changing the number of layers affects the overall accuracy.

Finetuned layer(s)	Accuracy
8	63 %
6,7 and 8	68.9 %
5,6,7, and 8	59 %

TABLE I: How number of layers impact finetuning

We wanted to explore how the number of images we train the neural network on impacts accuracy and how changing the learning rate, gamma, impacts accuracy as well [27]. Table II shows the test results we got on AlexNet. For all three cases, the layers we finetuned were 7 and 8.

Training Images	Testing Images	Gamma	Iterations	Accuracy
10, 703	861	0.1	20000	71.43%
10, 703	861	0.01	20000	72.94%

TABLE II: Finetuning AlexNet with larger set and different learning rate

Table III shows how sensitive the accuracy of the network is to size of the database. With only an increase of about 1,000 images to train the AlexNet Neural Network the accuracy increased by about 3 percent. For all runs, the layers we finetuned were 7 and 8.

Training Images	Testing Images	Gamma	Iterations	Accuracy
10, 703	861	0.1	20000	71.43%
11, 791	617	0.1	20000	74.59%

TABLE III: Finetuning with a larger set of images

2) GoogleNet Test Results: Table IV shows the GoogleNet test results. The only variable we changed with the GoogleNet test runs is the number of iterations. There was no change in accuracy, this is further discussed in the conclusion though.

Training Images	Testing Images	Gamma	Iterations	Accuracy
10, 703	861	0.1	4000	65.85%
10, 703	861	0.1	10000	65.85%
10, 703	861	0.1	20000	65.85%

TABLE IV: Finetuning with a larger set of images

VII. Conclusion and Future Work

Our results show a few important factors. When comparing our results to a research team's results from Stanford who conducted experimentation on identifying Skin Cancer Images, our data can be concluded to be accurate and fairly successful as well. The highest accuracy we obtained in identifying if the image of skin cancer is malignant or benign is 74.59

percent. The team at Stanford who conducted their research on detecting skin cancer with deep neural networks and compared it to what dermatologists would identify it as. Their neural network's results had an overall accuracy of " 72.1 0.9 percent (mean s.d.) " [10]. With such close percentages between our data and theirs we can safely say that our results prove to be accurate and correct to depend on. Our slightly higher percent is most likely due to training on a GPU since GPUs are faster and have more memory than CPUs. Taking the highest accuracy percent we achieved, we can say our network assures 74.59 percent of the skin cancer images it processes would be correctly identified as malignant or benign.

However, the two Software Engineers who developed their experimentation on using CNNs to diagnose Skin Cancer achieved an accuracy of 85 percent [9] thus we still have work to do in terms of what we can modify and test to gain even higher accuracy. That said, though, since our percentages are in the range of accuracy other research team's and engineer's diagnosing skin cancer with CNNs are getting, our results are promising to further expand the application of AlexNet in image recognition for medical diagnosis on health issues other than skin cancer. However, to be more secure with the accuracy and trustworthiness of AlexNet's ability to diagnosis medical images correctly, further experimentation would need to be conducted with a larger skin cancer database. Our 74.59 percent test result is based on 617 test images only, this limits how many correct and incorrect images it might classify - the more test images it runs on the more reliable its test results actually are. Going forward with next steps, since there are already studies focused on classifying skin cancer with Neural Networks, it is best to further test AlexNet on more Skin Cancer images so we have outside data to fact-check our results with when evaluating if the percentage of accuracy we get is actually accurate.

A second factor to conclude from our data, is a closer look at one of our experiments (Table I) that focused on the number of layers being trained and how they correlate to the final results. We assumed that the more layers that are trained the more accurate the results would be, until we trained all three fully connected layers plus a Convolutional Neural Network Layer and saw that the accuracy rate turned in the opposite direction. When we trained only 1 layer, layer 8 (a fully connected layer), we achieved 63 percent accuracy. When we trained 3 layers, layer 6, 7 and 8, there was an improvement of accuracy to 68.9 percent. And in the last case of training four layers, 5 (a CNN layer), 6, 7 and 8, the accuracy came out to be 59 percent.

Furthermore, data from Table II presents evidence that the learning rate impacts overall accuracy. Based on Table II, when the gamma (the learning rate) drops below 0.1 the accuracy increases. Most likely this is because with a smaller learning rate there is a smaller change in step size when the neural network is training the weights to adjust based on its previous iterations and thus is more sensitive to the image characteristics. However, dropping the learning rate below 0.01 could prove the opposite affect and increasing the learning rate above 0.1 could also prove to actually improve accuracy. But due to how long each test takes to run, technical issues we run to and time constraints, we were only able to manipulate each variable a few amount of times so we can test how changing different types of variables impacts a neural network's accuracy. Changing the learning rate is open for further investigation in future experiments to run.

Across all three of our data tables presented, another factor to conclude is a very evident factor in the results which is: the larger the database the neural network trains on the more accurate it is.

And lastly, between GoogleNet and AlexNet the Neural Network that proved to have higher accuracy is AlexNet. However, we would like to further investigate why the accuracy rate of the GoogleNet was not increasing proportionally with the number of iterations (as all previous tests on AlexNet have shown us) as well as why it is lower than AlexNet's accuracy since GoogleNet has more layers.

That being said, all the data we've derived gives us helpful information for to progress our research in future experimentation and training. Tweaking all the variables comes down to a balance, a "sweet spot" in regards to what number of layers, what learning rate, and what number of images and iterations will prove to result in the highest accuracy. Considering our goal is to provide as accurate results to any user, all these variables come into play in continuing further experimentation.

Both networks shows reliable results, the AlexNet more than the GoogleNet at this point in our research, that we can continue to fine tune and train on to work on improving accuracy performance. Once we have more data to understand the performance of deep learning image recognition and how different factors, such as the number of layers, learning rate, and database size impact accuracy in identifying skin cancer images we would like to expand our databases to train on medical images for other health issues, such as simply maybe a pink eye, a bruised ankle that's swollen, or even chickenpox The biggest weaknesses we face in working on research pertaining to image recognition, though, is finding data. It's an obstacle we've worked around and will continue to work around because without data we can't create technology to diagnose patients if we have nothing to go off of to teach a neural network how to make decisions. In finding the best architecture/parameters for a CNN as well as expanding what diseases CNNs can classify, will allow us to build onto the Vital Signs Multi-Sensor Kit [1] to develop a well-rounded tool that can cover a larger range of diseases for diagnosing which in turn will provide more medical care opportunities for our users. At that stage, we would reach to deliver a rounded medical care kit that anyone around the world could use, regardless of what local medical resources they have.

A. Acknowledgments

We would like to thank Santa Clara University for the support they have given us in pursuing our research, from resources, lab equipment and factulty support. Furthermore, we would like to thank Santa Clara University for the funding they provide us to pursue our research, and a special thank you to the Clare Boothe Luce Foundation for providing funding through Santa Clara University to pursue our research. We would furthermore like to thank our research professor, Professor Ogunfunmi, for the guidance, direction and support he has given us. This has been a great learning opportunity we are grateful to have undergone. We would also like to acknowledge Randol Spaulding for introducing us to the ISIC database based on similar research he has conducted. Lastly, we would like to thank Vineet Panchbhaiyye who is a Masters student in the Electrical Engineering department for being a teacher and mentor to us this past summer, he held weekly lectures to teach about the fundamentals of machine learning and introduced us to all the libraries we worked with to develop the scripts to modify the ISIC database as needed.

REFERENCES

[1] N. Arrizon, J. Hernandez, T. Ogunfunmi, A. Maldonado-Liu, A. Pacheco, and U. Kim, "Vital sensor kit for use with telemedicine in developing countries," in *2017 IEEE Global Humanitarian Technology Conference (GHTC)*, Oct 2017, pp. 1–5.

[2] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1153–1159, May 2016.

[3] D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck, "Deep Learning for Identifying Metastatic Breast Cancer," *ArXiv e-prints*, Jun. 2016.

[4] T. W. B. Data, "Data world bank." [Online]. Available: https://data.worldbank.org/indicator/SH.MED.PHYS.ZS

[5] W. H. Organization, "Early cancer diagnosis saves lives, cuts treatment costs," http://www.who.int/news-room/detail/03-02-2017-early-cancer-diagnosis-saves-lives-cuts-treatment-costs.

[6] A. I. for Cancer Research, "Global cancer statistics for the most common cancers," https://www.wcrf.org/dietandcancer/cancer-trends/worldwide-cancer-data.

[7] T. KUBOTA, "Deep learning algorithm does as well as dermatologists in identifying skin cancer," https://news.stanford.edu/2017/01/25/artificial-intelligence-used-identify-skin-cancer/.

[8] C. Y. W. N. D. Louisa G. Gordon, corresponding author Thomas M. Elliott and W. Visser, "Modelling the healthcare costs of skin cancer in South Africa," https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4818961/.

[9] J. R. Linik, "Skin Cancer Detection Using Artificial Intelligence," https://iq.intel.com/skin-cancer-detection-using-artificial-intelligence/.

[10] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, p. 115, 2017.

[11] "Artificial intelligence algorithms appear to be better at detecting skin cancer," https://www.news-medical.net/news/20180529/Artificial-intelligence-algorithms-appear-to-be-better-at-detecting-skin-cancer.aspx.

[12] "Man against machine: AI is better than dermatologists at diagnosing skin cancer," https://www.sciencedaily.com/releases/2018/05/180528190839.htm.

[13] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1207–1216, May 2016.

[14] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Commun. ACM*, vol. 521, no. 436, pp. 436–444, May 2015, nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved. [Online]. Available: https://www.nature.com/articles/nature14539

[15] G. Anthes, "Artificial intelligence poised to ride a new wave," *Commun. ACM*, vol. 60, no. 7, pp. 19–21, Jun. 2017. [Online]. Available: http://doi.acm.org/10.1145/3088342

[16] D. K. Gunnam, "Machine learning foundations -an intuitive approach," T. editor, Ed., IEEE Workshops on Machine Learning and Convolutional Neural Networks. The address of the publisher: IEEE, 7 2017, p. 110, an optional note.

[17] S. University, "Stanford alexnet slides." [Online]. Available: http://vision.stanford.edu/teaching/cs231b_spring1415/slides/alexnet_tugce_kyunghee.pdf

[18] I. Lutkebohle, "Kratzert Finetuning AlexNet," 2008, [Online; accessed 19-July-2008]. [Online]. Available: http://kratzert.github.io/2017/02/24/finetuning-alexnet-with-tensorflow.html

[19] I. Ltkebohle, "Kaggle Cats Vs. Dogs," 2008, [Online; accessed 19-July-2008]. [Online]. Available: https://www.kaggle.com/c/dogs-vs-cats-redux-kernels-edition/data

[20] A. Moujahid, "A Practical Introduction to Deep Learning with Caffe and Python," http://adilmoujahid.com/posts/2016/06/introduction-deep-learning-python-caffe/.

[21] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[22] L. A. dos Santos, "Artificial Intelligence," https://leonardoaraujosantos.gitbooks.io/artificial-inteligence/content/googlenet.html.

[23] Kaggle, "Cats and Dogs Images," https://www.kaggle.com/c/dogs-vs-cats-redux-kernels-edition/data.

[24] I. S. I. Collaboration, "ISIC," https://isic-archive.com/.

[25] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. S. Bernstein, A. C. Berg, and F. Li, "Imagenet large scale visual recognition challenge," *CoRR*, vol. abs/1409.0575, 2014. [Online]. Available: http://arxiv.org/abs/1409.0575

[26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[27] J. Czaja, "Solver Prototxt," https://github.com/intel/caffe/wiki/Solver-Prototxt.