International Journal of Social Science and Economic Research

ISSN: 2455-8834

Volume:06, Issue:10 "October 2021"

A NOVEL META-MACHINE LEARNING PLATFORM ABLE TO AUTONOMOUSLY LEARN HOW TO DIAGNOSE AUTISM, BREAST CANCER, MELANOMA MOLE CANCER, AND PINK EYE

Nihal Boina, Jai Agarwal, Taruna Agarwal, Chloe Song, Vihaan Cherukuri, Smaran Pasupulati, Sathvik Redrouthu, Kunal Saxena, Jishnu Patel, Sri Vaishnavi Konagalla, Jay Pallepati, Rashmith Repala, Ashana Patel, Sudhit Sangela, Samarth Jain, Kinnari Chaubal, Veda Kalwala, Aneesh Sreedhara, Nihar Xavier, Yash Sonis, Anjali Sanapala, Sachin Nimmalapudi, Tanya Singhal, Avyakt Gaur, Poorvaja Gopinath, Zaakir Khan, Nikhit Rachapudi, Saketh Chintalapati, Karthik Yella, Riya Pasupulati, Zaynah Khan, Ishan Doma, Avesta Zahedi, Saranya Ganne, Aastha Sharma, Dhruv Khatod, Malachai Onwona, John Leddo

MyEdMaster, LLC, 13750 Sunrise Valley Drive, Herndon, VA, United States of America

John Leddo is director of research at MyEdmaster.

DOI: 10.46609/IJSSER.2021.v06i10.041 URL: https://doi.org/10.46609/IJSSER.2021.v06i10.041

ABSTRACT

One of the main goals of machine learning is to make a General Artificial Intelligence. Currently, human artificial intelligence researchers work on meticulously manipulating model parameters by hand in order to arrive at highly optimized machine learning models. In the future, a system will be needed such that a software is able to completely arrive at an optimized model to a specific topic all by itself. In this study, a novel machine learning platform was created that could learn how to diagnose a variety of diseases including autism, breast cancer, melanoma mole cancer, and pink eye without being explicitly taught to learn them. Artificial intelligence (AI) algorithms, specifically convoluted neural networks (CNN), have been employed to diagnose these skin diseases. Additionally, these algorithms were manipulated by an automated hyperparameter manipulator, using machine learning to find, sort, and train, validate, and test on a dataset all by itself. Put simply, we were able to make software capable of making its own algorithms through a meta-machine learning approach, aiming to fill the role of an AI researcher. Additionally, the software was able to achieve consistent overall testing accuracy of at least 90%, quantifying its potential use in diagnosing fitting diseases that it was not explicitly taught to learn in the first place.

Keywords: artificial intelligence, convolutional neural network, deep learning, autism, breast cancer, melanoma, pink eye

Introduction

Autism refers to a broad range of conditions characterized by challenges with social skills, repetitive behaviors, speech and nonverbal communication. According to the Centers for Disease Control, autism affects an estimated 1 in 54 children in the United States today.

Autism is a spectrum disorder, meaning that each person with autism had a distinct set of strengths and challenges. Signs of autism usually appear by age 2 or 3. Some associated development delays can appear even earlier, and often, it can be diagnosed as early as 18 months. Additionally, some research suggests that it is caused by a combination of genetic and nongenetic or environmental influences, providing the possibility that autism could be predicted from just a visual image. There has still yet to be a popularised machine learning model to accurately identify autism.

Breast cancer is a tumor that grows in the breast when cells in the breast grow out of control. Ultrasound, mammograms, MRIs, and biopsies are commonly used to diagnose this disease, and these methods are not entirely accurate, with Mammograms being 87% in identifying breast cancer and MRIs being 90% accurate. There has still yet to be a popularised machine learning model to accurately identify breast cancer.

Melanoma is a version of skin cancer that forms when the melanin in the skin (the cells that give the skin its tan or brown color) starts to “grow out of control”. Other names for this cancer include malignant melanoma and cutaneous melanoma. Cancerous melanoma can appear on the skin as small moles or marks that start to grow and evolve. It is more dangerous than other skin cancers because it is more likely to rapidly spread. In the United States for example, less than 200,000 living cases of melanoma exist. For identifying a “mole” symptom for melanoma, physicians generally look at typical exposed places on the body, such as the neck, arms, legs, and hands. There has still yet to be a popularised machine learning model to accurately identify melanoma.

Pink eye is a very common condition that is caused by the eyes coming into contact with viruses, allergies, or bacterias. This condition is usually noted when the sclera (the white visible portion of the eyeball) is a pink/red color. Patients with pink eye experience discharge and a stinginess in the eye. Pink eye usually lasts up to a week or two, and has no damage to the eyes . Additionally, this disease is almost always diagnosed by people, and there has not been any popularised machine learning model at the moment to model it.

ISSN: 2455-8834

Volume:06, Issue:10 "October 2021"

Traditionally, for all of these diseases, people have to use time and energy to carefully examine the possibility of these diseases in various subjects, which AI can help vastly in doing more accurately and consistently. Since the early 1950's, work has been done in artificial intelligence (AI), but there have been many limiting factors furthering AI into the medical field. With the current advent of AI in the medical field, AI assists doctors in detecting meaningful relationships between datasets; this leads to clinical diagnoses or suggestions in treatments. The introduction of deep learning has made AI in medicine a possibility. One of the first applications of AI in medicine was the development of a consultation program for glaucoma that was created at Rutgers University. In addition, systems that recommend antibiotics based on the physician's knowledge and a system that provides possible diagnoses based on imputed symptoms have been created. In recent times, AI has vastly seen applications in gastroenterology, endoscopy, diagnosis of cancers, and many other advantageous applications. Artificial Intelligence (AI) and CNN algorithms are being heavily employed in diagnosis of several cancers, including breast and lung cancer. The ability of the algorithm to comb through thousands and even millions of pieces of data gives it the appeal to apply to the medical field.

Convoluted neural networks (CNN) take input as an image dataset and filter these images through a series of layers (convoluted, pooling, and fully connected). The algorithm will then apply specified weights to features of images to apply it to the classification probability distribution. The first layers identify basic features of the images such as pixel lines (horizontal and vertical) through examination of brightness and darkness of pixels, and angles.

Furthermore, an aspect of CNNs that make them especially appealing to image classification in the medical field is its ability to surmise important patterns in the visual that will help create a probability distribution for the classifications. This unique asset is possible by the convolution and pooling of the pixel values for the image. By learning and relearning the data, applying weights to features within segments of the images, the algorithm identifies the discrepancies within an image of a dataset and is able to hone in on the minute differences that separate different categories.

However, there is one problem with conventional machine learning methods: that the dataset is usually hand-picked by machine learning researchers without careful finetuning of the model's parameters through human interaction. Imagine this scenario: if you ask software to detect for skin cancer, it should search up a dataset, download it, and train and recurse through various types of models to finally arrive at the best model to use directly and potentially save for future use. By doing this, more intelligence is given to the software and the potential for creating one software that can ubiquitously model various diseases all by itself is realized.

ISSN: 2455-8834

Volume:06, Issue:10 "October 2021"

There are three steps in a machine learning project: the acquisition of data, pre-processing, and model finetuning, fitting, and testing. In this paper, we discuss a technology that can systematically generate accurate models with select pre-processing all by itself to arrive at highly optimized machine learning algorithms for diagnosing autism, breast cancer, melanoma mole cancer, and pink eye by itself provided user input.

Literature Review

Due to the rapid advances of machine learning, machine learning has been applied more frequently in the medical field. There have been recent studies concentrated on the use of machine learning to diagnose skin diseases. Bhadula, S. et al found that through comparing various Machine Learning Classifiers, a CNN was observed to fit labelling acne the best, with an error rate of 0.04 [6]. The power of CNNs in image classification has been quantified in numerous studies, which is why the type of model used in this study was constantly a CNN.

The one problem that many machine learning projects run into is the lack of images for training and the accuracy rate of these algorithms. By training the software with a small data-set, the algorithm usually becomes over-fitted, in which it merely memorizes the data fed into it instead of learning patterns from it. Overfitting also occurs when the algorithm is run at an excessively high epoch level. This is why a large dataset with a low epoch level is favored by the software's data acquisition and training sequence.

Another potential limitation of machine learning is that data often need to be supplied to the machine learning algorithm in order for it to train itself on the data. This limitation is benign in cases where the task being learned is relatively static, such as learning to read numbers entered on a check being deposited: there are only 10 different digits to learn and there may be limitations in the reasonable variations of each digit. Similarly, this limitation may be benign in cases where the task being learned is dynamic but the data are well organized and automatically supplied to the algorithm so that it may update its model, e.g., analyzing customer purchasing decisions made on a website.

However, not all domains are well structured. Knowledge is constantly changing as are people's demands and expectations for technology. Imagine a future society in which the Internet of Things (IoT) is highly integrated into the fabric of our devices. In such a world, routine bathroom objects such as mirrors, toothbrushes, hairbrushes, electric razors, etc. may perform routine medical diagnoses as we use them. Equipped with cameras, chips and connections to the Internet, these devices may monitor our bodies and process images and other data regarding our health.

ISSN: 2455-8834

Volume:06, Issue:10 "October 2021"

Medical knowledge, however, evolves rapidly. Such knowledge should not be hard-coded into the system as that would risk the system becoming outdated. Moreover, people may expect the system to be "all-knowing" and ask it a variety of medical questions, even those the system may not be initially programmed to answer. Ideally, then, the system should be evolving and dynamic. It should respond as a human would when confronted with a question requiring an answer s/he does not already know: it should search for relevant information, learn the information and then respond to the question posed.

The present paper describes a system designed to do just that. Users pose questions like "do I have acne?" and the system searches the Internet for relevant data, learns the data and then processes a picture of the user to answer the question.

The Current System

The current system is designed to respond to user queries of autism, breast cancer, melanoma mole cancer and pink eye. The software had to utilize several imports using python to function, all listed below

1. Os
2. Tensorflow
3. Shutil
4. Image from PIL
5. Open datasets
6. Selenium
7. SciPy
8. Pandas

Technology

The software starts out by first accepting user input as to what kind of disease to check for. In this case, the diseases to check for were "autism","breast cancer", "melanoma mole cancer", and "pink eye". Upon doing this, the software then goes to the popular machine learning website kaggle.com, and searches for a related dataset through a selenium python import. The software does this by entering the disease name and specifying whether to look for datasets on the Kaggle page. Following this, the software then downloads the dataset all by itself using the open datasets python import.

Once the dataset is retrieved, the software is then able to try a variety of data preprocessing methods and machine learning algorithms to finally arrive at a unique highly optimized model that it can use for the specific user-inputted disease. In this case, the inquiries were "autism","breast cancer", "melanoma mole cancer", and "pink eye", leading the software to use the internet and select the following datasets.

A descriptive flowchart of this process is shown in Figure 1.

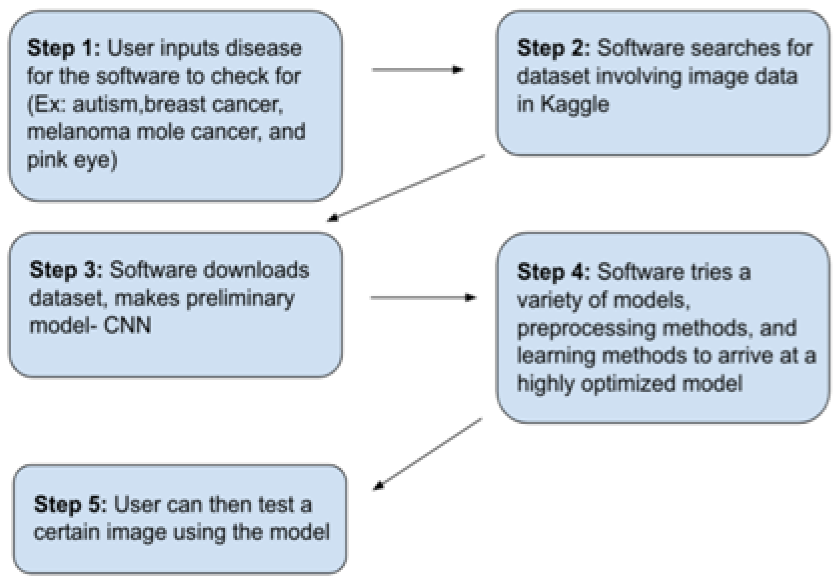

Figure 1: Flowchart of Processes of Software

Dataset:

The research leveraged four independent datasets: for autism, breast cancer, melanoma mole cancer, and pink eye. All of these leveraged datasets consisted of image type data.

The first dataset consisted of 2000 facial images of children that did and did not have autism. The dataset was split up such that 1000 images were children that were autistic and the other 1000 were of children that were not autsitic. This dataset was acquired from the popular machine learning dataset website kaggle.com. Table 1 shows the image breakdown for autism and non-autism for this dataset. Figure 2 shows a sample autism and non-autism picture.

ISSN: 2455-8834

Volume:06, Issue:10 "October 2021"

Category	Number of Images
Autism	1000
Non-Autism	1000

Table 1: Image File Breakdown for Autism Dataset

Figure 2: Sample Pictures Autism Dataset

The second dataset consisted of 1578 breast ultrasound images of benign, malignant, and normal cases of breast cancer. The dataset was split up such that 891 images were benign, 421 images were malignant, and the remaining 266 were normal. This dataset was acquired from the popular machine learning dataset website kaggle.com. Table 2 shows the breakdown of images from this dataset and Figure 3 shows sample images.

ISSN: 2455-8834

Volume:06, Issue:10 "October 2021"

Category	Number of Images
Benign	891
Malignant	421
Normal	266

Table 2: Image File Breakdown for Breast Cancer Dataset

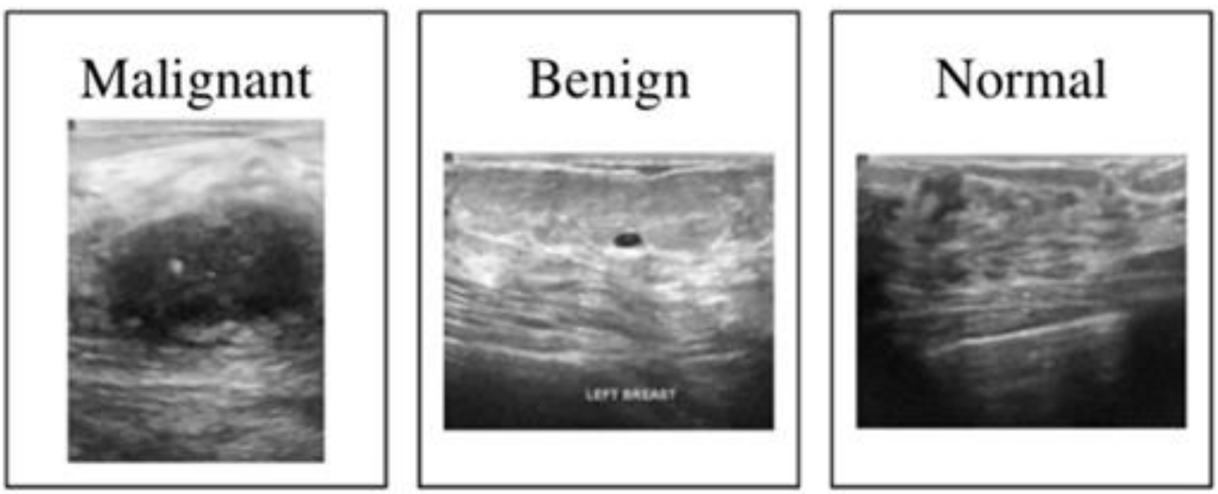

Figure 3: Sample Pictures Breast Cancer Dataset

The third dataset consisted of 3597 images of benign, and malignant cases of skin cancer. The dataset was split up such that 1800 images were benign and 1497 were normal. This dataset was acquired from the popular machine learning dataset website kaggle.com . Table 3 shows the image breakdown for this dataset and Figure 4 shows sample images.

International Journal of Social Science and Economic Research

ISSN: 2455-8834

Volume:06, Issue:10 "October 2021"

Category	Number of Images
Benign	1800
Malignant	1497

Table 3: Image File Breakdown for Skin Cancer

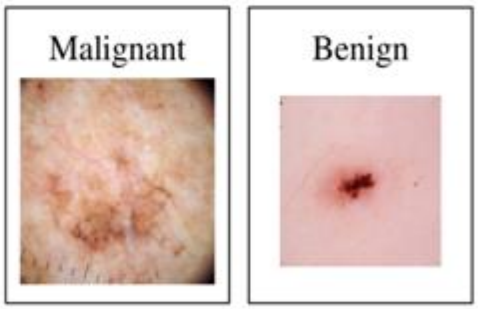

Figure 4: Sample Pictures Skin Cancer Dataset

The fourth dataset consisted of a thousand images of eyes that were considered healthy and pink. The dataset was split up such that 500 images were normal and 500 were pink eye. This dataset was acquired from google images under the search term 'Pink Eye'. Table 4 shows the image breakdown for this dataset and Figure 5 shows sample images.

Category	Number of Images
Benign	500
Malignant	500

Table 4: Image File Breakdown for Pink Eye

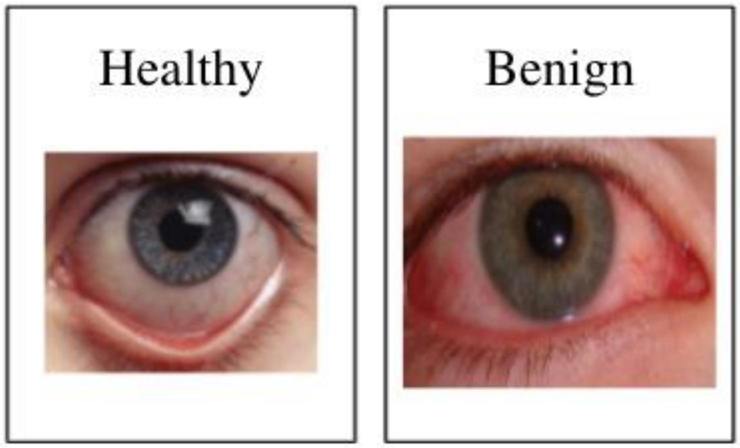

Figure 5: Sample Pictures Pink Eye Dataset

Algorithm:

A Convolutional Neural Network algorithm was trained in each to identify whether a picture of a person's face exhibited the certain disease. This specific algorithm processes data that exhibits a natural shift-invariance in the images which then detects minor differences. In each dataset, the software was automated to start off with default parameters declaring a CNN with a fixed input layer, with four hidden layers using Adam and ReLU, and various parameters were then changed with each iteration to achieve the best randomized testing results with each trained model.

Model Construction Procedure:

In order to arrive at the best possible model, the software goes through a variety of different parameters while training on 80% of the total data. There was no validation data utilized in this paper. The training data and testing data was split in a 4:1 ratio. After each subsequent testing accuracy, the software alternates the physical parameters in order to achieve better testing accuracy. The physical parameters that software changes about the model and preprocessing methods is the inputted size of the image, the number of nodes in each subsequent layer, the number of layers (max: 10), the activation functions, grayscaling, gaussian smoothing, and the addition of an autoencoder.

Tools and Technology:

ISSN: 2455-8834

Volume:06, Issue:10 "October 2021"

We have leveraged the GPU (graphics processing unit) as the hardware accelerator with HIGH-RAM runtime shape. Python 3.7 was used to write the code for the algorithm. The model was trained on Google Colab and used the default Keras and Tensorflow packages.

Results

Table 5 shows that each model produced a high testing accuracy for each according dataset, specifically 93.7%, 97.5%, 95.4%, and 99.4%, all of which were above a conventional 90%.

Topic	Final Testing Accuracy
Autism	93.70%
Breast Cancer	97.50%
Skin Cancer	95.40%
Pink Eye	99.40%

Table 5: Final Testing Accuracies for Both Models

Conclusion

This software is able to accurately learn how to diagnose autism, breast cancer, mole cancer, and pink eye. With the optimized epoch values, inclusion of gaussian smoothing and image resizing, manipulation of other hyperparameter values, as well as the development of other neural layers with optimal conditions, this software is able to learn to diagnose select diseases based on user input with accuracies that all exceed 90%.

What makes this software unique is its ability to accept user input, search relevant training data on its own, train itself and then perform at exceptionally high levels. In other words, the software accepts a user task and then acts completely autonomously to train itself and then perform the task the user specified. As we think to the future where devices are inter-connected

ISSN: 2455-8834

Volume:06, Issue:10 "October 2021"

through the Internet of Things, virtually unlimited information will be available to these devices, thus giving them unlimited possibilities to serve their users.

Currently, when we think of commercial AI and machine learning applications, they are limited to information retrieval (search engines or personal assistants) or learning narrowly defined skills based on specified datasets. The former places huge burdens on their users to sift through the retrieved information and learn it. The latter results in systems that are highly specialized, e.g., reading checks deposited at an ATM or analyzing purchasing behaviors. In the future, we want our systems to be flexible and powerful: able to accept virtually unlimited requests from users and then search for and learn relevant information on their own, so that they can perform their users' tasks. The present project represents such a first step in the medical space. In Leddo and Liang [6], we describe a similar process for learning mathematics. More research is needed in this area.

References

[1]What is autism? (n.d.). Autism Speaks. Retrieved October 8, 2021, from https://www.autismspeaks.org/what-autism

[2]CDC. (2020, March 13). Screening and diagnosis | autism spectrum disorder (Asd) | ncbddd. Centers for Disease Control and Prevention. https://www.cdc.gov/ncbddd/autism/screening.html

[3]CDCBreastCancer. (2021, September 22). What is breast cancer screening? Centers for Disease Control and Prevention. https://www.cdc.gov/cancer/breast/basic_info/screening.htm

[4]Melanoma of the skin statistics | cdc. (2021, June 8). https://www.cdc.gov/cancer/skin/statistics/index.htm

[5]Reed, K. B., Brewer, J. D., Lohse, C. M., Bringe, K. E., Pruitt, C. N., & Gibson, L. E. (2012). Increasing incidence of melanoma among young adults: An epidemiological study in olmsted county, minnesota. Mayo Clinic Proceedings, 87(4), 328–334. https://doi.org/10.1016/j.mayocp.2012.01.010

[6] Machine learning algorithms based skin disease detection. (n.d.). ResearchGate. Retrieved September 8, 2021, from https://www.researchgate.net/publication/341372376_Machine_Learning_Algorithms_based_Skin_Disease_Detection

[7]Kaggle: Your machine learning and data science community. (n.d.). Retrieved September 8, 2021, from https://www.kaggle.com

ISSN: 2455-8834

Volume:06, Issue:10 "October 2021"

[8]CDC. (2021, August 5). Protect yourself from pink eye. Centers for Disease Control and Prevention. http://www.cdc.gov/conjunctivitis/

[9]Red eye: Causes, symptoms, treatments, prevention. (n.d.). Cleveland Clinic. Retrieved October 8, 2021, from https://my.clevelandclinic.org/health/symptoms/17690-red-eye

[10]Weber, G. (n.d.). *Autistic Children Dataset* [Data set]. Kaggle. https://www.kaggle.com/gpiosenka/autistic-children-data-set-traintestvalidate

[11]Shah, A. (2021). Breast Ultrasound Images Dataset [Data set]. Kaggle. https://www.kaggle.com/aryashah2k/breast-ultrasound-images-dataset

[12]Fanconi, C. (2019). Skin Cancer: Malignant vs. Benign [Data set]. Kaggle. https://www.kaggle.com/fanconic/skin-cancer-malignant-vs-benign

[13] Leddo, L. and Liang, I. (2021). Incorporating Abstract Knowledge Structures in Machine Learning: Improving Question Answering, Problem Solving and Teaching in Personal Assistants and Educational Software. *International Journal of Social Science and Economic Research, 6(2)*, 661-673.

