*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 ACNE AND JAUNDICE

Nihal Boina, Jai Agarwal, Taruna Agarwal, Ishita Samant, Nikhil Doma, Ishaan Saxena, Parth Patel, Ashwin Menachery, Bhargav Vantikommu, Kunal Singhal, Sai Varun Konagalla, Akshara Sanapala, Reema Patel, Sophia Nasibdar, Deepika Ravi, Sathvik Lakamsani, Mason Elkas, John Leddo

John Leddo is director of research at MyEdmaster.

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

DOI: 10.46609/IJSSER.2021.v06i10.040 URL: https://doi.org/10.46609/IJSSER.2021.v06i10.040

## ABSTRACT

Acne is a skin condition caused by hair follicles that become clogged with oil and dead skin cells. Around 85% of people have experienced acne at some time in their lives, and acne is known to have heavy emotional and physical impacts on regular people. Jaundice is another type of skin condition that is noted by the yellow pigmentation of skin and whites of the eyes due to high bilirubin levels. Jaundice is rare in adults, but is significantly more common in newborn babies, with around 80% affected by it during the first week of life. Artificial intelligence (AI) algorithms, specifically convoluted neural networks (CNN), have been employed to diagnose these skin diseases. Additionally, these algorithms would be manipulated by an automated hyperparameter manipulator, using extensive machine learning to find, sort, and train, validate, and test on a dataset all by itself. Put simply, we were able to make an automated software capable of making its own state-of-the-art algorithms through a meta-machine learning approach, filling 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 skin diseases and fitting diseases that it was not explicitly taught to learn in the first place.

**Keywords:** artificial intelligence, convolutional neural network, acne, deep learning, jaundice

## Introduction

Acne is the most common skin condition and occurs when hair follicles become clogged with oil, dead skin cells, or bacteria and is most common in teens but affects people of all ages [1].

  

ISSN: 2455-8834

Volume:06, Issue:10 "October 2021"

Roughly 3 in every 4 people aged 11 to 30 years are affected by acne. There are different levels of acne, usually dependent on a person's genetics and lifestyle. Depending on the severity, acne can cause emotional distress and also scar the skin permanently [2]. Acne can be fixed with ointment and a change in diet. Currently, any person can easily diagnose acne, but there has still yet to be a uniform way for a person to use software to get an accurate representation of the degree of acne present and potential solutions.

Jaundice is a condition noted by the yellowish pigmentation of the skin and whites of the eyes [3]. Jaundice is rare in adults but is fairly common in babies, with 80% of new-borns affected during the first week of their life [4]. The primary cause of jaundice is excess bilirubin, but it can also be caused by internal bleeding (hemorrhage), an infection in the blood (sepsis), and liver malfunction. Severe jaundice comes with the risk of bilirubin passing into the brain and causing significant damage to the baby [5]. Many hospitals have policies for examining babies for jaundice, but there has still yet to be a machine learning model to accurately identify jaundice.

For both acne and jaundice, 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

  

ISSN: 2455-8834

Volume:06, Issue:10 "October 2021"

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 home 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 a person asks software to detect for acne, 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.

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 acne and jaundice all by itself.

**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 acne and other 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.

**Software Components:**

The software has to utilize several imports using python to function, all listed below

1. Os

  

ISSN: 2455-8834

Volume:06, Issue:10 "October 2021"

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 acne and jaundice. Upon doing this, the software will then go to the popular machine learning website kaggle.com, and search for a related dataset through a selenium python import [7]. The software does this by entering the disease name and specifying whether to look for datasets on the Kaggle page. Following this, the software will then download 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 pre-processing 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 "acne" and "jaundice", leading the software to use the internet and select the following datasets.

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

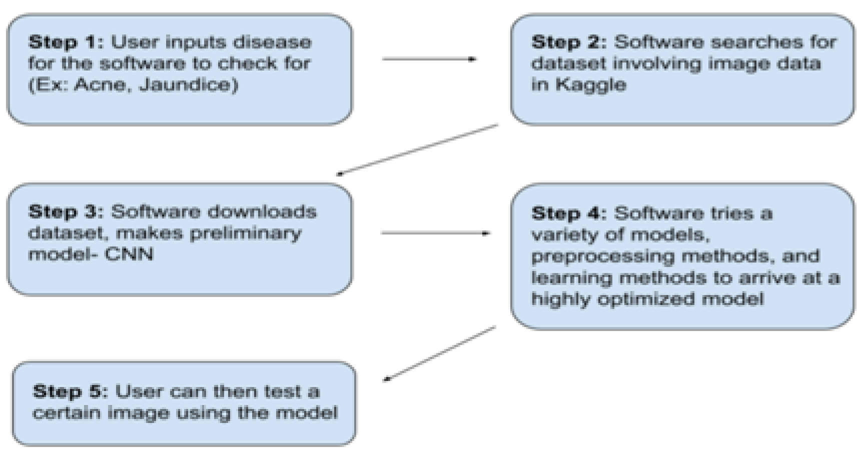


**Figure 1: Flowchart of Processes of Software**

  

ISSN: 2455-8834

Volume:06, Issue:10 "October 2021"

**Dataset:**

The research leveraged two independent datasets: one for the acne images and the other for the jaundice pictures. The first dataset consisted of 999 facial images of three separate acne severities: minimal, moderate, excessive/dangerous acquired from the popular machine learning dataset website kaggle.com [7]. Table 1 shows the breakdown of the minimal/no, moderate and excessive images in the acne dataset and Figure 2 shows image samples of each level of acne.

| Category | Number of Images |
|---|---|
| Minimal/No | 387 |
| Moderate | 473 |
| Excessive | 139 |

**Table 1: Image File Breakdown for Acne Dataset**

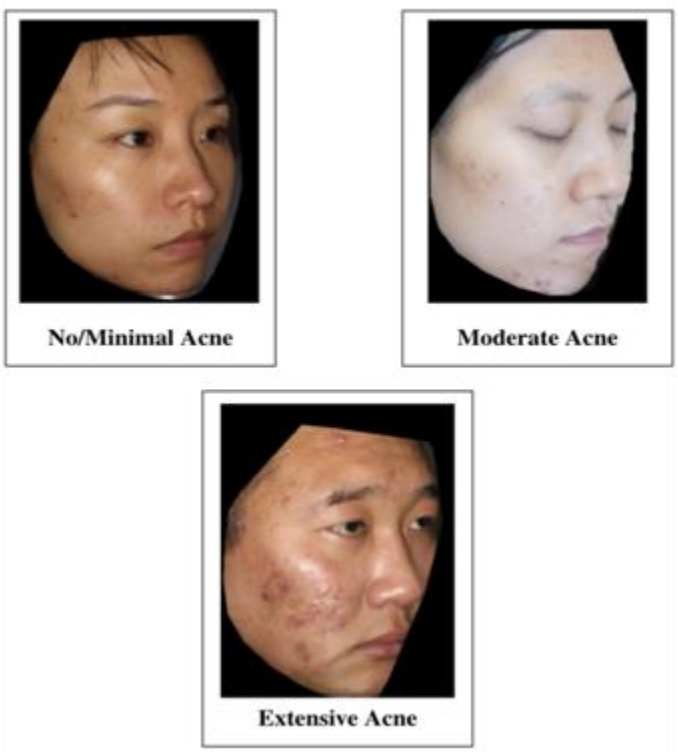


**Figure 2: Sample Pictures Acne Dataset**

  

ISSN: 2455-8834

Volume:06, Issue:10 "October 2021"

The second dataset consisted of 18 images of babies with or without jaundice. Table 2 shows the breakdown of the jaundice/without jaundice images in the jaundice dataset and Figure 3 shows image samples of each level of jaundice.

| Category | Number of Images |
|---|---|
| Jaundice | 13 |
| Without Jaundice | 6 |

**Table 2: Image File Breakdown for Jaundice Dataset**


**Figure 3: Sample Pictures Jaundice Dataset**

**Algorithm:**

A Convolutional Neural Network (CNN) 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

  

ISSN: 2455-8834

Volume:06, Issue:10 "October 2021"

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 constant 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:**

We have leveraged the GPU (graphics processing unit) as the hardware accelerator with a HIGH-RAM runtime shape. Python 3.7 was used to write the code for the algorithm. The model was trained on Pycharm and used the most up to date Keras and Tensorflow packages as of 10/08/21.

**Results**

Table 3 shows that each model produced a high testing accuracy for each dataset, specifically 93.5% for jaundice and 97.40% for acne.

| Topic | Final Testing Accuracy |
|---|---|
| Jaundice | 93.50% |
| Acne | 97.40% |

**Table 3: Final Testing Accuracies for Both Models**

**Conclusion**

Our results show that high testing accuracies were obtained by both models, both of which were above a standard 90%. Thus, it is supported that the software is able to accurately learn how to diagnose acne and jaundice. With the optimized epoch values, inclusion of gaussian smoothing

  

ISSN: 2455-8834

Volume:06, Issue:10 "October 2021"

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 select diseases based on user input. The next step would be then to first implement diseases outside of acne and jaundice involving solely image data with the software and then expanding the model to take various forms of input. Eventually, the software will be expanded to provide generated forms of treatment, in which more research will have to be conducted in order to execute.

**References**

[1] Nancy Garrick, D. D. (2016, September 1). Acne. National Institute of Arthritis and Musculoskeletal and Skin Diseases. https://www.niams.nih.gov/health-topics/acne

[2] Acne—Symptoms and causes. (n.d.). Mayo Clinic. Retrieved September 8, 2021, from https://www.mayoclinic.org/diseases-conditions/acne/symptoms-causes/syc-20368047

[3]Jaundice. (n.d.). [Text]. Retrieved September 8, 2021, from https://medlineplus.gov/jaundice.html

[4]Kaplan, M., & Hammerman, C. (2017). 97—Hereditary contribution to neonatal hyperbilirubinemia. In R. A. Polin, S. H. Abman, D. H. Rowitch, W. E. Benitz, & W. W. Fox (Eds.), Fetal and Neonatal Physiology (Fifth Edition) (pp. 933-942.e3). Elsevier. https://doi.org/10.1016/B978-0-323-35214-7.00097-4

[5]Infant jaundice—Symptoms and causes. (n.d.). Mayo Clinic. Retrieved September 8, 2021, from https://www.mayoclinic.org/diseases-conditions/infant-jaundice/symptoms-causes/syc-20373865

[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