Application of Auto-Keras to COVID-19 Chest X-Ray Image Classification

CPT_S 570, Fall 2020 Final Project

Thunradee Tangsupakij
Computer Science Department
Washington State University
Pullman, USA
t.tangsupakij@wsu.edu

***Abstract*—In this paper, we apply the Auto-Keras algorithm to classifying COVID-19 chest x-ray images. Although the method performs better than our baseline models, we find the performance to be poor relative to other modern approaches. Our finding along with the limitation of our approach warrants further consideration to Auto-Keras and possibly SVM for COVID-19 diagnosis.**

I. Introduction

The 2019 Corona VIrus Disease (COVID-19) is a severe respiratory-based virus affecting millions of people across the globe. As of December 2020, there has been over 15 million cases claiming over 288 thousand lives in the United States alone [1]. Outbreak of the virus is characterized by rapid spread through airborne moisture droplets, whereby one may become infected by inhaling some of the droplets [2]. A necessity in containing the spread of COVID-19 involves early detection of infection in an individual. This allows the individual who has the virus to quarantine, and if necessary, seek emergency medical attention. Efforts towards fast detection of infection has led to the widely-available Rapid Diagnostic Test (RDT) [3], which have been found to be moderately effective at accurately diagnosing COVID-19 [4]. Despite the rapid testing capabilities of RDT, precise diagnostic tests are needed in hospital settings for patients experiencing severe symptoms.

Newly admitted patients into emergency medical facilities who are experiencing chest pain or difficulty breathing frequently undergo chest x-ray imaging in order to evaluate the problem. Other forms of evaluation, including Reverse Transcription Polymerase Chain Reaction (RT-PCR) tests and Computed Tomography (CT) scans are rarely immediately available in most emergency facilities. Often-times these patients have not been tested for COVID-19 prior to admittance; the next step in treating the patient often involves employing COVID-19 tests where both the waiting time for feedback from the test and the accuracy of the test are imperative to proper treatment of the patient. A study conducted with 64 patients in [5] reported 69% diagnostic accuracy from professionals using x-ray imaging alone.

II. Related Work

The dual necessity for expeditious and highly accurate diagnostic testing for COVID-19 has led to the application of machine learning (in particular, deep learning) to the problem, where learning algorithms may be employed to classify x-ray images taken from patients as containing COVID-19 or not containing COVID-19. A deep-learning approach to this problem was initiated in [6], where the authors applied a Convolution Neural Network (CNN) over 128 x-ray images. Their approach garnered an f-measure of .95, greatly outperforming the results found in [5]. A more involved technique was employed in [7], where the authors developed a two-stage algorithm. In the first stage, features were extracted from multiple convolutional neural networks using correlation-based feature selection [8]. In the second stage, a BayesNet classifier utilizing the features extracted from the first stage classified each x-ray as having COVID-19 or not. This methodology was applied to 560 x-ray images leading to an accuracy of 91.16% and to an f-measure of .914. A third model example was applied in [9] involving an ensemble of three popular neural-network architectures: DenseNet, ResNet, and VGGNet. Unlike the previous papers mentioned, this approach included training the three neural-network architectures over a much larger data-set including 15,959 x-ray images. The results of this model garnered an f-measure of .946, the highest of all the above models mentioned. Further deep-learning approaches to model the COVID-19 outbreak not limited to x-ray imaging include [10]–[13].

III. Methods

The approaches discussed in the last section showed relatively strong performance in predicting COVID-19 relative to the 69% accuracy found in [5] for professional diagnostics. The models trained in these approaches involve artificial neural networks under the deep learning framework, a notoriously computationally expensive framework where these models may demand orders of magnitude greater computational power in order to train [14]. Furthermore, a broad range of network architectures have been developed with varying degrees of success in a number of fields; for a general introduction to

deep learning with neural networks, we refer the reader to [15].

Neural Architecture Search (NAS), a sub-field of the broader Automated Machine Learning (AutoML) paradigm is a relatively new idea with the goal of finding the best neural network architecture for a specified problem. Following the notation of [16], by best we mean a neural network h^* that satisfies:

$$h^* = argmin_{f \in \mathcal{F}} Cost(f(\theta^*), D_{val}), \tag{1}$$

$$\theta^* = argmin_{\theta} \mathcal{L}(f(\theta), D_{train}), \tag{2}$$

where $Cost(.,.)$ is the evaluation metric and θ^* is the learned parameter. The search space, denoted here as $\mathcal{F}$ denotes the architectures to be searched over and that are obtainable from the initial architecture. Although these ideas directly address the issue of finding the best neural-network architecture, they are generally very computationally expensive [16], and may require an extremely large dataset in order to be accurate. The most popular of these approaches have been gradient based methods [17], [18], evolutionary algorithms [19], and reinforcement learning [20]–[22]. The major drawback most of these approaches suffer from is the need to train each new neural network completely from scratch, hence the need for extremely large data sets and massive computational power.

The goal of this paper is to test the recently introduced Auto-Keras NAS algorithm [16] on classifying COVID-19 in x-ray images. Auto-Keras extends the above ideas on NAS algorithms by implementing network morphism [23]. The idea here is simple: rather than re-train each new archecture from scratch, one can morph, or alter the current architecture in small ways, for example by adding a layer. To conduct this morphism, a kernalized distance metric defined in [16] establishes the kernalized distance $\kappa(f_a, f_b)$ (thought of as the amount of morphing needed to change one network to the other) as:

$$\kappa(f_a, f_b) = exp(-\rho^2(d(f_a, f_b))), \tag{3}$$

where $d(f_a, f_b)$ is the edit-distance between two networks.

Guiding this network morphism is a Bayesian optimization routine [25]. This routine can be thought of as conducting three steps: first, it updates the neural network by morphing, followed by a generation step where the updated network under it's new architecture is created. Finally, the last step is the observation step where the newly created network is trained on the data. For details with the full mathematical description of this process, we refer the reader to [16], where the authors first introduced auto-keras. In conducting this experiment, we will train and test the Auto-Keras formulation and compare it against two baseline methods: Logistic Regression and Support Vector Machine (SVM) with Radial Basis Function (RBF) kernel. For the SVM used in this experiment, we conduct a grid-search for the parameter values using the following set of values:

- **RBF kernel** C: [0.01, 0.1, 1, 10, 100], γ: [0.010, 001, 0.0001]

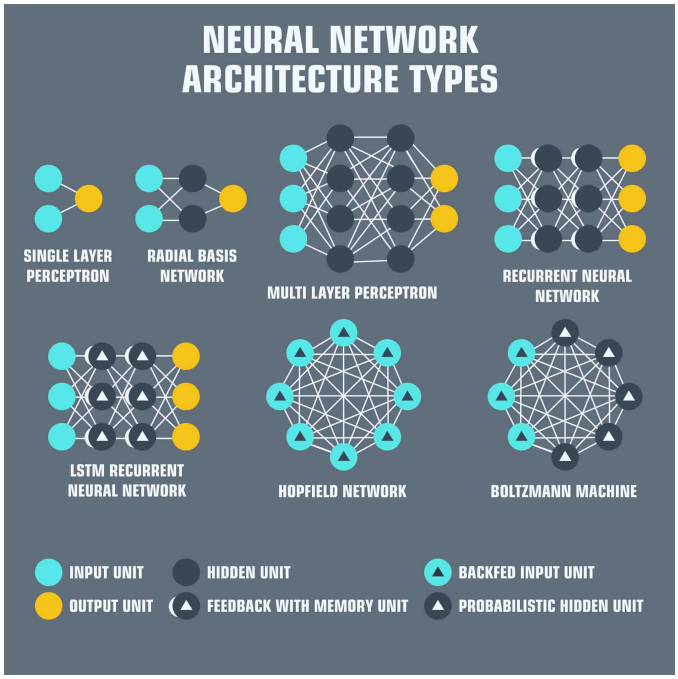


Fig. 1: Example Neural-Network Architectures [24].

A. Dataset and Preprocessing

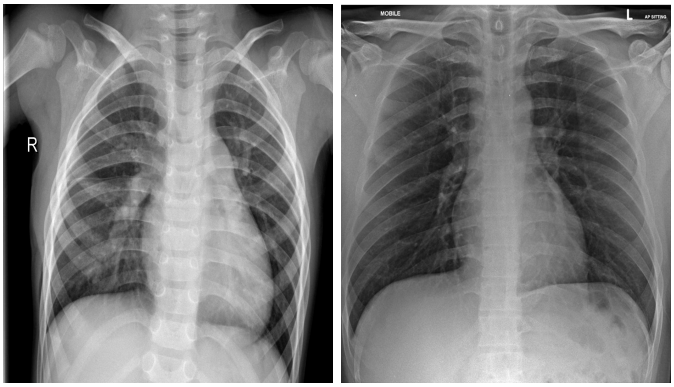
We started this experiment by downloading and organizing the sample images into suitable folders. We then resized the images to 250x250 pixels and normalized to 0-1 scale. The images were originally RGB format, so for Logistic Regression and SVM classifiers we converted the images to grey scale. We included 563 COVID-Pneumonia and 260 other viral and bacterial Pneumonia images from [26], [27]. 71 COVID X-ray images from [28]. 374 other viral and bacterial Pneumonia images from [29]. In total, the combined datasets contains 634 COVID-19 X-ray images and 634 Non-COVID-19 X-ray images. We divided the dataset into a training set and testing set; 80% dedicated to the training set and 20% to the testing set. An example x-ray image of a patient containing COVID-19 and an example x-ray image of a patient containing Pneumonia caused by bacteria is contained in Fig.2.

B. Evaluation Criteria

To measure the effectiveness of Auto-Keras on our dataset, we chose to follow the example from set out in the previous studies in [7], [8], [10], [11] by measuring the accuracy, precision, recall and f-measure statistics.

- **Accuracy**: This simple measure takes the percentage of correctly identified cases out of the entire dataset. The higher the accuracy, the better. In general, accuracy is considered a poor measure of model effectiveness in cases where the classes are imbalanced. It is computed as follows:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{4}$$



(a) Pneumonia caused by bacteria chest x-ray [29].

(b) Pneumonia caused by COVID-19 chest x-ray [26], [27]

Fig. 2: Example x-ray images

```
Model: "functional_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_1 (InputLayer)         [(None, 250, 250, 3)]     0
_________________________________________________________________
cast_to_float32 (CastToFloat (None, 250, 250, 3)       0
_________________________________________________________________
normalization (Normalization (None, 250, 250, 3)       7
_________________________________________________________________
separable_conv2d (SeparableC (None, 248, 248, 32)      155
_________________________________________________________________
separable_conv2d_1 (Separabl (None, 246, 246, 64)      2400
_________________________________________________________________
dropout (Dropout)            (None, 246, 246, 64)      0
_________________________________________________________________
flatten (Flatten)            (None, 3873024)           0
_________________________________________________________________
dropout_1 (Dropout)          (None, 3873024)           0
_________________________________________________________________
dense (Dense)                (None, 1)                 3873025
_________________________________________________________________
classification_head_1 (Activ (None, 1)                 0
=================================================================
Total params: 3,875,587
Trainable params: 3,875,580
Non-trainable params: 7
_________________________________________________________________
```

Fig. 3: Model summary of Auto-Keras result

- **Precision**: Calculated as the percentage of true positives of COVID-19 divided by the total number of predicted positives. It is computed as follows:

$$Precision = TP/(TP + FP) \tag{5}$$

- **Recall**: This metric is a measure of completeness, which is calculated as the percentage of positives that were correctly identified as true positives divided by the number of actual positives. It is computed as follows:

$$Recall = TP/(TP + FN) \tag{6}$$

- **F-measure**: This is a combination of precision and recall that provides a significant measure for a test dataset that includes an imbalanced class. It is computed as follows:

$$F - measure = 2\left(\frac{Precision \times Recall}{Precision + Recall}\right) \tag{7}$$

IV. Results

The output from running AutoKeras is displayed in Fig.3. This output characterizes the architecture constructed by the algorithm; we refer the reader to [16] for more details on output interpretation.

	Clasifiers		
	LR	SVM	AutoKeras
Accuracy	0.77	0.85	0.88
Precision	0.77	0.81	0.88
Recall	0.75	0.92	0.89
F-Measure	0.76	0.86	0.88

TABLE I: Accuracy, precision, recall and f-measure by classifier

We found the worst performing model of this study to be the Logistic Regression, which has an accuracy of 0.77 , precision of 0.77 , recall of 0.75 , and an f-measure of 0.76. The SVM with the optimized parameters (C=10, gamma=0.0001, kernel=rbf) performed relatively well compared to Auto-Keras, where the SVM has accuracy of 0.85, precision of 0.81, recall of 0.92 and f-measure of 0.86. In comparison to these base models, Auto-Keras performed better on all of our evaluation criteria except recall. Our Auto-Keras was found to have an accuracy of 0.88, precision of 0.88, recall of 0.89 and f-measure of 0.88.

V. Conclusion and Future Directions

Machine learning and more specifically, deep learning algorithms have become powerful modern tools for predicting and solving a variety of complex problems. The emergence of COVID-19 has been a devastating world-wide virus, with over 1.6 million deaths on record. Application of deep learning models for classifying x-ray images of patient lungs who are suspected to have COVID-19 has led to faster and more accurate diagnoses of COVID-19, thus allowing these patients to get the care they need while possibly preventing a patient with COVID-19 spreading the virus any further.

Previous studies on classifying and predicting COVID-19 from x-ray images using machine learning techniques primarily used generalizations of a few popular neural network architectures, especially convolution-based architectures. In this study, we implemented Auto-Keras, an automated artificial neural network architecture search algorithm designed to not only find the best network architecture for a problem, but also optimize its parameters. The results of this study were primarily aimed at measuring its effectiveness relative to two baseline models: logistic regression and SVM with RBF kernel. Our results indicate that Auto-Keras is capable of outperforming both logistic regression and SVM with RBF kernel on nearly all evaluation criteria. The only criteria it failed to outperform in was recall, where we found the SVM with RBF kernel performed marginally better. However, due to the classification imbalance prevalent in many x-ray chest imaging datasets, our evaluation criteria was the f-measure, where we found Auto-Keras outperformed the baseline models.

A secondary goal of this study was to compare the training time of Auto-Keras to our baseline approaches. Despite the alleged advantages Auto-Keras has over alternative methods in computational complexity, we found the model was difficult

to run on our standard machines. Furthermore, we found installation of a compiler using the GPU on a computer to be challenging, so we oped to use Google Colab. However, this led to crashes due to memory limits (despite using the pro version). We were forced to set Auto-Keras with max trials = 30 and max epochs = 100. If we could overcome these issues and find a way to allow the model to run over more epochs, there is a strong possibility we would have found a better performing model. For the outputted model, it took our machine two hours and 37 minutes to train the model, in comparison to only 21.8 seconds for the logistic regression and one hour, 24 minutes for the SVM with RBF kernel. This indicates that for roughly twice the computational complexity, Auto-Keras is capable of small to moderate gains in capability relative to SVM with RBF kernel.

We believe our study provides a strong basis for further study. In particular, we believe a thorough testing of the Auto-Keras algorithm with allowance for higher-numbers of epochs could provide a stronger idea of what the model is capable of. Furthermore, our tests indicate that SVM-based methods may be applicable to the problem, especially in cases where computational complexity may be an issue when training these models. In the end, these models indicate strong ability to diagnose COVID-19 using x-ray imaging alone, warranting further reasearch in this area.

Code is available here:

- https://github.com/Thunradee/COVID-19-Chest-X-Ray-Image-Classification

References

[1] *Cdc covid data tracker*. [Online]. Available: https://covid.cdc.gov/covid-data-tracker/.

[2] L. Morawska and J. Cao, "Airborne transmission of sars-cov-2: The world should face the reality," *Environment International*, p. 105 730, 2020.

[3] *Simple / rapid tests*. [Online]. Available: https://www.who.int/news-room/q-a-detail/simple-rapid-tests.

[4] D. O. Andrey, P. Cohen, B. Meyer, G. Torriani, S. Yerly, L. Mazza, A. Calame, I. Arm-Vernez, I. Guessous, S. Stringhini, *et al.*, "Diagnostic accuracy of augurix covid-19 igg serology rapid test," *European journal of clinical investigation*, vol. 50, no. 10, e13357, 2020.

[5] H. Y. F. Wong, H. Y. S. Lam, A. H.-T. Fong, S. T. Leung, T. W.-Y. Chin, C. S. Y. Lo, M. M.-S. Lui, J. C. Y. Lee, K. W.-H. Chiu, T. Chung, *et al.*, "Frequency and distribution of chest radiographic findings in covid-19 positive patients," *Radiology*, p. 201 160, 2020.

[6] M. Alazab, A. Awajan, A. Mesleh, A. Abraham, V. Jatana, and S. Alhyari, "Covid-19 prediction and detection using deep learning," *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 12, pp. 168–181, 2020.

[7] B. Abraham and M. S. Nair, "Computer-aided detection of covid-19 from x-ray images using multi-cnn and bayesnet classifier," *Biocybernetics and biomedical engineering*, vol. 40, no. 4, pp. 1436–1445, 2020.

[8] M. A. Hall, "Correlation-based feature selection for machine learning," 1999.

[9] M. Karim, T. Döhmen, D. Rebholz-Schuhmann, S. Decker, M. Cochez, O. Beyan, *et al.*, "Deepcovidexplainer: Explainable covid-19 predictions based on chest x-ray images," *arXiv preprint arXiv:2004.04582*, 2020.

[10] J. P. Cohen, L. Dao, P. Morrison, K. Roth, Y. Bengio, B. Shen, A. Abbasi, M. Hoshmand-Kochi, M. Ghassemi, H. Li, *et al.*, "Predicting covid-19 pneumonia severity on chest x-ray with deep learning," *arXiv preprint arXiv:2005.11856*, 2020.

[11] R. R. Bouckaert, "Bayesian network classifiers in weka for version 3-5-7," *Artificial Intelligence Tools*, vol. 11, no. 3, pp. 369–387, 2008.

[12] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. J. Soufi, "Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning," *arXiv preprint arXiv:2004.09363*, 2020.

[13] L. Wang, Z. Q. Lin, and A. Wong, "Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images," *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020.

[14] D. Justus, J. Brennan, S. Bonner, and A. S. McGough, "Predicting the computational cost of deep learning models," in *2018 IEEE International Conference on Big Data (Big Data)*, IEEE, 2018, pp. 3873–3882.

[15] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*, 2. MIT press Cambridge, 2016, vol. 1.

[16] H. Jin, Q. Song, and X. Hu, "Auto-keras: An efficient neural architecture search system," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 1946–1956.

[17] H. Cai, L. Zhu, and S. Han, "Proxylessnas: Direct neural architecture search on target task and hardware," *arXiv preprint arXiv:1812.00332*, 2018.

[18] H. Liu, K. Simonyan, and Y. Yang, "Darts: Differentiable architecture search," *arXiv preprint arXiv:1806.09055*, 2018.

[19] T. Desell, "Large scale evolution of convolutional neural networks using volunteer computing," in *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, 2017, pp. 127–128.

[20] B. Baker, O. Gupta, N. Naik, and R. Raskar, "Designing neural network architectures using reinforcement learning," *arXiv preprint arXiv:1611.02167*, 2016.

[21] H. Pham, M. Y. Guan, B. Zoph, Q. V. Le, and J. Dean, "Efficient neural architecture search via parameter sharing," *arXiv preprint arXiv:1802.03268*, 2018.

[22] S. Xie, H. Zheng, C. Liu, and L. Lin, "Snas: Stochastic neural architecture search," *arXiv preprint arXiv:1812.09926*, 2018.

[23] H. Cai, T. Chen, W. Zhang, Y. Yu, and J. Wang, "Efficient architecture search by network transformation," *arXiv preprint arXiv:1707.04873*, 2017.

[24] Joshinav, *3 types of neural networks that ai uses: Artificial intelligence*, Apr. 2019. [Online]. Available: https://www.allerin.com/blog/3-types-of-neural-networks-that-ai-uses.

[25] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," *Advances in neural information processing systems*, vol. 25, pp. 2951–2959, 2012.

[26] J. P. Cohen, P. Morrison, and L. Dao, "Covid-19 image data collection," *arXiv 2003.11597*, 2020. [Online]. Available: https://github.com/ieee8023/covid-chestxray-dataset.

[27] J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, and M. Ghassemi, "Covid-19 image data collection: Prospective predictions are the future," *arXiv 2006.11988*, 2020. [Online]. Available: https://github.com/ieee8023/covid-chestxray-dataset.

[28] Larxel, *Covid-19 x rays*, Mar. 2020. [Online]. Available: https://www.kaggle.com/andrewmvd/convid19-x-rays?select=xrays.csv..

[29] P. Mooney, *Chest x-ray images (pneumonia)*, Mar. 2018. [Online]. Available: https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia.