EXPLAINABLE-BY-DESIGN APPROACH FOR COVID-19 CLASSIFICATION VIA CT-SCAN

A PREPRINT

Plamen Angelov *
School of Computing and Communications
LIRA Research Centre
Lancaster University
Lancaster, LA1 4WA, UK
p.angelov@lancaster.ac.uk

Eduardo Soares
Lancaster University
Lancaster, LA1 4WA, UK
e.almeidasoares@lancaster.ac.uk

April 24, 2020

ABSTRACT

The COVID-19 disease has widely spread all over the world since the beginning of 2020. On January 30, 2020 the World Health Organization (WHO) declared a global health emergency. At the time of writing this paper the number of infected about 2 million people worldwide and took over 125,000 lives, the advanced public health systems of European countries as well as of USA were overwhelmed. In this paper, we propose an eXplainable Deep Learning approach to detect COVID-19 from computer tomography (CT) - Scan images. The rapid detection of any COVID-19 case is of supreme importance to ensure timely treatment. From a public health perspective, rapid patient isolation is also extremely important to curtail the rapid spread of the disease. From this point of view the proposed method offers an easy to use and understand tool to the front-line medics. It is of huge importance not only the statistical accuracy and other measures, but also the ability to understand and interpret how the decision was made. The results demonstrate that the proposed approach is able to surpass the other published results which were using standard Deep Neural Network in terms of performance. Moreover, it produce highly interpretable results which may be helpful for the early detection of the disease by specialists.

Keywords Interpretability · Explainable-by-design approach · Human interpretable rules · Prototype-based · COVID-19

1 Introduction

In December 2019, an outbreak coronavirus (SARS-CoV-2) infection began in Wuhan, the capital of central China's Hubei province [1, 2, 3]. On January 30, 2020 the World Health Organization (WHO) declared a global health emergency [4] and with some delay and hesitation on 11 March 2020 WHO declared pandemic. By 14 April 2020, accumulative 1,985,135 confirmed cases and 125,344 deaths were documented [5]. USA has become the new epicenter of the disease with 605,354 documented cases and 25,394 deaths (14 April 2020) [5].

Researchers of different disciplines work along with public health officials to understand the COVID-19 pathogenesis and jointly with the policymakers urgently develop strategies to control the spread of this new disease [6]. Recent findings have observed imaging patterns on chest radiography and computed tomography (CT) for patients diagnosed with COVID-19 [7, 8, 9, 10].

Prospective analysis revealed bilateral lung opacities on 40 of 41 (98%) chest CTs in infected patients in Wuhan and described lobular and subsegmental areas of consolidation as the most typical findings [6]. Other investigators found high rates of ground-glass opacities and consolidation, sometimes with a rounded morphology and peripheral lung distribution [11, 10]. Thoracic radiology evaluation is often key to the evaluation of patients suspected of COVID-19

*Honorary Professor, Technical University, Sofia, Bulgaria.

infection [12]. Prompt detection and diagnosis of the disease is invaluable in the efforts to ensure timely treatment. From a public health perspective, rapid patient isolation is crucial for containment of this communicable disease[4] and optimal use of available resources which quickly become scarce and overwhelmed by the exponentially growing number of patients and prolonged periods of treatment.

Recently, artificial intelligence (AI) and, specifically, deep learning based approaches have demonstrated high levels of performance in the medical imaging domain due to their ability to automatically extract latent features and bypass so called "*handcrafting*" [13]. In addition, the technique called *transfer learning* [14] made possible to train a deep neural network on one set of images (e.g. ImageNet [15]) but use effectively on another set of images. While manual reading CT and X-ray images takes 15 minutes and involves a highly skilled medical doctor/consultant which are now in high demand the use of AI and deep learning can take few seconds on a computer and be automated which provides opportunity for high throughput and remote way of operation. However, few stumbling blocks still hamper the wider use of deep neural networks. These include: i) their opaque, "black-box" nature and inability to explain any decision [16]; ii) their inability to continue to learn once trained, to learn from a handful of examples and data and compute power appetite [17].

In this paper we present a new deep learning method that is explainable by design. It is able to continue to learn and adapt for each new data sample which is immensely important for the case of COVID-19 (and other disease), because new cases are being accumulated every minute and traditional approaches require either iterative re-training or ignores the new data. The proposed approach is non-iterative and is entirely based on recursive calculations and use of prototypes. Therefore, it is computationally very efficient. The architecture of the proposed method combines reasoning and learning in a synergy while alternative approaches focus on either reasoning which favours the interpretability and explainability or on machine learning and statistical approaches which favour the accuracy and other statistical measures for the expence of the interpretability and explainability. In this paper we demosntrate that the proposed approach can be very efficient in detecting COVID-19 via CT scans and can be very useful to explain the decisions which itself may also be very important for medical doctors.

The main idea of the proposed approach is based on prototypes (images of CT scans - both, with and without COVID-19) and is using the density in the data/feature space to build empirical estimations of the distributions [18]. The proposed approach is non-iterative and non-parametric, which explains its efficiency in terms of time and computational resources. From the user perspective, the proposed approach is clearly understandable/explainable which for the specific case of COVID-19 infections means to aid explanations and decisions made ultimately by human doctors (instead of percentages and likelihoods they can see and understand an image and compare similarities). In this sense, this is an approach of anthropomorphic machine learning [18]. We tested the proposed method on the very recent COVID-CT-Dataset [19] which contains a set of real cases. Results have demonstrated that proposed approach provides superior performance's measured by F1 score and other metrics, but also, critically, it offers epxlainability and is able to continue to learn from new data.

2 Methods and Algorithms

2.1 Concept and Basic Algorithm

Same as most machine learning methods, the proposed in this paper method starts with pre-processing which involves scaling, augmentation, and rotation. In order to extract features of the CT images we use transfer learning over the GoogleNet Deep Learning structure [14]. It is important to stress that GoogleNet is used just to define the feature space and it was not trained on the CT images, but on ImageNet [15]. Other approaches could also be used for this purpose.

The prototype-based learning is the core of the proposed method (Fig. (1)). The prototypes are actual training data samples (in this case, images) which are highly representative (local peaks of the density and empirically derived probability distributions [18]). They are focal points of locally valid generative models described by multi-modal Cauchy distribution [18].

The algorithm of the proposed approach is described below. With the first observed image (data sample) it is being converted to a vector of features using transfer learning. In this paper, we use a vector with size 1000 formed from the last fully connected layer of the GoogleNet [14]. More information about the pre-processing step for the proposed method can be found in the Supplementary Material available for this paper.

Let $T = \{(x_i, c_i)\}_{i=1}^{N}$ be training data set with $x_i \in \mathbb{R}^n$ denoting the feature vector and $c_i \in \{1, 2\}$ denoting the class (COVID-19 or No COVID-19) for each $i \in \{1, ..., N\}$. N is the number of training data/images used.

The proposed algorithm works per class; therefore, all the calculations are done for each class separately.

The meta-parameters are initialized with the first observed data sample.

$$\mu \leftarrow \bar{x}_1; \quad ; \mathrm{V}^1 \leftarrow \{\bar{x}_1\}; \quad p^1 \leftarrow \bar{x}_1; \quad S^1 \leftarrow 1; \quad P \leftarrow 1; \quad r^1 \leftarrow r_o \tag{1}$$

where μ denotes the mean; V_1 denotes the first cluster; p_1 is the first prototype of the first cluster, V_1; S_1 is the corresponding support (number of members); P is the total number of the identified prototypes; r_1 is the corresponding radius of the area of influence of V_1(in this paper, we use $r^* = \sqrt{2 - 2cos(30^o)}$ same as [18]; the rationale is that two vectors for which the angle between them is less than $\pi/6$ or 30^o are pointing in close/similar directions. That is, we consider that two feature vectors can be considered to be similar if the angle between them is smaller than 30 degrees. Note that r^* is data derived, not a problem- or user- specific parameter. In fact, it can be defined without *prior* knowledge of the specific problem or data).

The next step is to calculate the data density at the current data point, $\bar{x}_i; i \in \{1, ..., N\}$.

$$D(\bar{x}_i) = \frac{1}{1 + \frac{||\bar{x}_i - \mu||^2}{\sigma^2}}; \tag{2}$$

Starting from the mutual distances (Euclidean or Mahalanobis type) between the data points (samples) in the feature space it can be demonstrated theoretically [18] that the data density takes the form of a Cauchy type function as in Eq. (2).

Then the algorithm absorbs the new data samples/images, $\bar{x}_i$ one by one by assigning then to the nearest (in the feature space) prototype, p^{j^*}:

$$j^* = \underset{j \in \{1,...,P\}}{\operatorname{argmin}} \{||\bar{x}_i - p^j||^2\} \tag{3}$$

Because of this form of assignment, the shape of the data partitioning is of the so-called Voronoi tessellation type [20]. We call all data points associated with a prototype *data clouds*, because their shape is not regular (e.g., hyper-spherical, hyper-ellipsoidal, etc.) and the prototype is not necessarily the statistical and geometric mean [18].

Then, using the density and the distance to the nearest prototype we check the following conditions [18] based on which we determine if the current data sample/image is going to be added to the set of prototypes as a new prototype or not:

$$\begin{aligned} &\textit{IF } (D(\bar{x}_i) \geq \max_{j \in \{1,...,P\}} (D(p^j))) \quad \textit{OR } (D(\bar{x}_i) \leq \min_{j \in \{1,...,P\}} (D(p^j))) \quad \textit{OR } (||p^{j^*} - \bar{x}_i|| > r^{j^*}) \\ &\qquad\qquad \textit{THEN } (\textit{add a new data cloud}) \end{aligned} \tag{4}$$

When adding a new data cloud the following updates are being made:

$$P \leftarrow P + 1; \quad \mathrm{V}^P \leftarrow \{\bar{x}_i\}; \quad p^P \leftarrow \bar{x}_i; \quad S^P \leftarrow 1; \; r^P \leftarrow r_o; \tag{5}$$

Alternatively, the meta parameters of the nearest data cloud are being updated as follows [18]:

$$\begin{aligned} V^{j^*} &\leftarrow V^{j^*} + \{\bar{x}_i\}; \\ p^{j^*} &\leftarrow \frac{S^{j^*}}{S^{j^*} + 1} p^{j^*} + \frac{S^{j^*}}{S^{j^*} + 1} \bar{x}_i; \\ S^{j^*} &\leftarrow S^{j^*} + 1; \\ r^{j^*} &\leftarrow \sqrt{\frac{(r^{j^*})^2 + (1 - ||p^{j^*}||^2)}{2}}; \end{aligned} \tag{6}$$

One of the strongest aspects of the proposed approach is its high level of interpretability which comes from its prototype-based nature. Linguistic *IF...THEN* expressions that represent human reasoning can be formed around the local generative models:

$$\mathrm{R}: \quad \textit{IF } (Image \sim p^1) \; \textit{OR} \; \ldots \; \textit{OR} \; (Image \sim p^P) \; \textit{THEN } (\textit{Class } c) \tag{7}$$

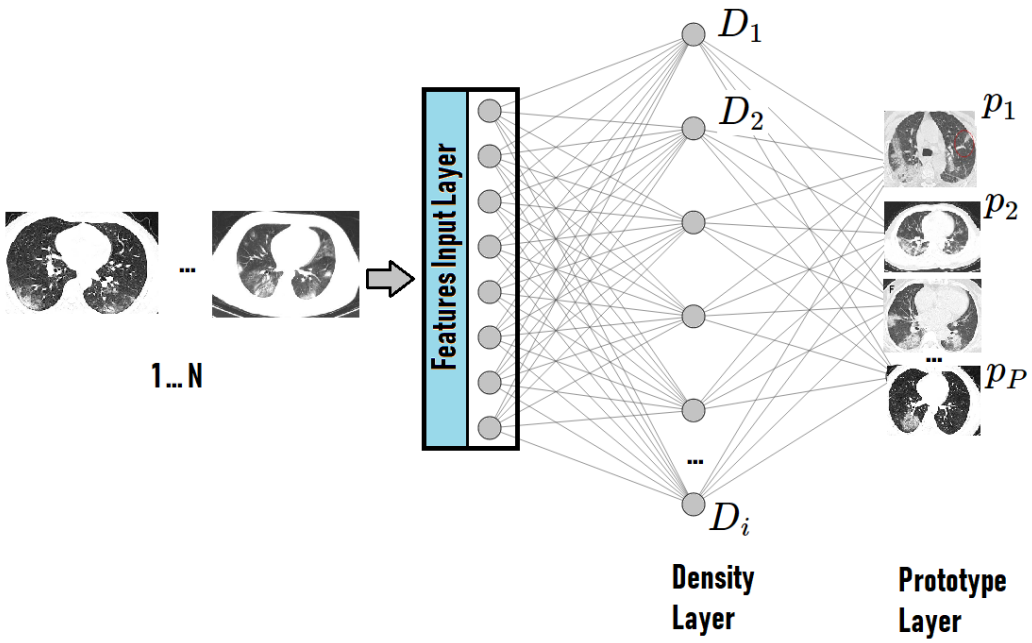


Figure 1: This figure illustrates the layered architecture of the proposed method. It has the form of a deep neural network but is using clear to understand prototypes (actual images). The density layer identifies the local peaks of the density and empirically derived probability distributions. The prototypes are actual training data samples (in this case, images) which are highly representative (local peaks of the density and empirically derived probability distributions.

The learning procedure of the proposed approach is summarized by the following algorithm.

Learning Procedure

Read the first feature vector sample x_i representing the image I_i of the class c;
Set $\mu \leftarrow \bar{x}_1$; ; $\mathrm{V}^1 \leftarrow \{\bar{x}_1\}$; $p^1 \leftarrow \bar{x}_1$; $S^1 \leftarrow 1$; $P \leftarrow 1$; $r^1 \leftarrow r_o$;
FOR i = 2, ...
  Read $\bar{x}_i$;
  Calculate $D(\bar{x}_i)$ and $D(p^j)$ $(j = 1, 2, ..., P)$ according to equation (2);
  IF Eq. (4) holds
    Create rule according to Eq. (5);
  ELSE
    Search for p^j according to Eq. (3);
    Update rule according to Eq. (6);
  END
END

3 Results

In this section we report the results obtained by the proposed eXplainable Deep Learning classification approach when applied to the COVID-CT-Dataset [19]. Results presented in Table 1 compare the proposed algorithm with other state-of-the-art approaches, including traditional (*black-box*) deep neural network, Support vector Machines, etc. In summary, the advantages of the proposed method include:

- high precision as compared with the top state-of-the-art algorithms.
- high level of explainabilit.
- no user- or problem- specific algorithmic meta parameters
- non-iterative algorithm able to learn continuously.

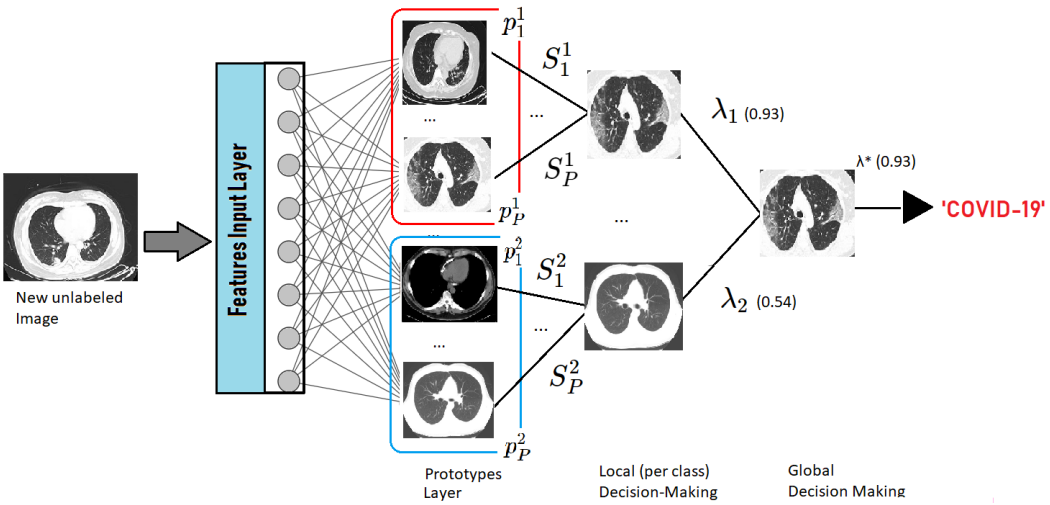


Figure 2: The figure illustrates the decision-making process of the proposed approach. As a new unlabeled data/image arrives it is compared to the identified prototypes. The local (per class) decision-making process is dedicated to calculate the winning prototype per class (the most similar to the new image prototypes of a given class). The global decision-making layer is in charge of forming the overall decision by comparing the degrees of similarity to all classes (in this case, to COVID-19 or NO COVID-19).

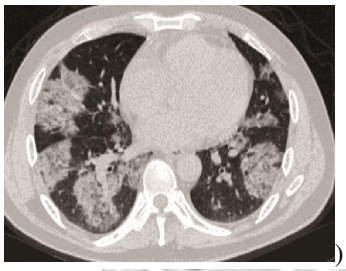
Method \ Metric	Accuracy	Precision	Recall	F1 Score	AUC
This paper	**88.6%**	89.7%	**88.6%**	**89.2%**	**88.6%**
Baseline [19]	84.%	**97.0%**	76.2%	85.3%	82.4%
SVM	80.5%	84.4%	83.5%	84%	79.7%
KNN	83.9%	90.4%	82.4%	86.2%	84.3%
AdaBoost	83.9%	87.7%	83.5%	85.5%	84%
Naive Bayes	70.5%	77%	73.6%	75.3%	69.6%

Table 1: The proposed eXplainable Deep Learning classifier provided better results in terms of Accuracy, Recall, $F1$ Score, and AUC. Besides the best results, the proposed approach also provided highly interpretable results that may be helpful for specialists. The proposed classifier identified 30 prototypes for non-COVID and 33 prototypes for COVID partients. Rules generated by the identified prototypes for COVID and Non-COVID patients are illustrated by Figs. (3) and (4) respectively. The baseline approach [19] is a Neural Network Deep Leaning-based approach which is 'black box' and the interpretability is very cost.

Using the proposed method we generated (extracted form the data) linguistic *IF...THEN* rules which involve actual images of both cases (COVID-19 and NO COVID-19) as illustrated in Figs. (3) and (4). Such transparent rules can be used in the decision-making process for early diagnostics for COVID-19 infection. Rapid detection with high sensitivity of viral infection may allow better control of the viral spread. Early diagnosis of COVID-19 is crucial for the disease treatment and control.

Fig. (5) illustrates the evolving nature of the proposed approach. The proposed approach is able to continuously learn as new data is presented to the system. Therefore, no full retraining is required due to its life-long learning architecture. In the opposite way, the Baseline approach [19] is based on Neural Network Deep Learning and requires full retraining for new data samples, what can be really cost in terms of time and computational complexity.

Computing tomography is a quick non-invasive imaging modality with high accuracy. According to [8, 9] almost all patients with COVID-19 had characteristic CT features during the disease, effects such as different degrees of ground-glass opacities with or without crazy-paving sign, multifocal organizing pneumonia, and architectural distortion in a peripheral distribution. The proposed approach has demonstrated high efficiency on the identification and classification of such characteristics, and then provide high accurate and interpretable results.

R: IF ($Image \sim$ ) OR ($Image \sim$ )
OR ($Image \sim$ ) OR ($Image \sim$ )
... OR ($Image \sim$ 

) THEN 'Infection by COVID-19'

Figure 3: Final rule given by the proposed eXplainable Deep Learning classifier for the COVID-19 identification. Differently from 'black box' approaches as deep neural networks, the proposed approach provides highly interpretable rules which can be used by human experts for the early evaluation of patients suspected of COVID-19 infection.

4 Conclusion

In this paper we present a new explainable deep learning approach for COVID-19 detection via CT Scan. The proposed approach demonstrates better results in terms of performance than other state-of-the-art approaches, surpassing the baseline Deep Neural Network approach in terms of performance. Moreover, it also provides epxlanations in the form of *IF...THEN* rules using actual images of CT scans with and without COVID-19. This is of great importance for medical specialists to understand and diagnose COVID-19 at early stages via computed tomography. In addition, this method is fast and can continue to learn from new images which is very important in a real life application. CT can accurately reflect the disease evolution and monitor the treatment effects [21]. Rapid detection and diagnostics of the disease is of supreme importance to ensure timely treatment, and rapid patient isolation in order to slow the spread of the disease [22].

In conclusion, chest CT imaging has high sensitivity for diagnosis of COVID-19. We offer a highly transparent deep learning approach which outperforms state-of-the-art approaches in order to detect COVID-19 via CT.

References

[1] Chaolin Huang, Yeming Wang, Xingwang Li, Lili Ren, Jianping Zhao, Yi Hu, Li Zhang, Guohui Fan, Jiuyang Xu, Xiaoying Gu, et al. Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. *The Lancet*, 395(10223):497–506, 2020.

[2] Na Zhu, Dingyu Zhang, Wenling Wang, Xingwang Li, Bo Yang, Jingdong Song, Xiang Zhao, Baoying Huang, Weifeng Shi, Roujian Lu, et al. A novel coronavirus from patients with pneumonia in china, 2019. *New England Journal of Medicine*, 2020.

[3] Fei Zhou, Ting Yu, Ronghui Du, Guohui Fan, Ying Liu, Zhibo Liu, Jie Xiang, Yeming Wang, Bin Song, Xiaoying Gu, et al. Clinical course and risk factors for mortality of adult inpatients with covid-19 in wuhan, china: a

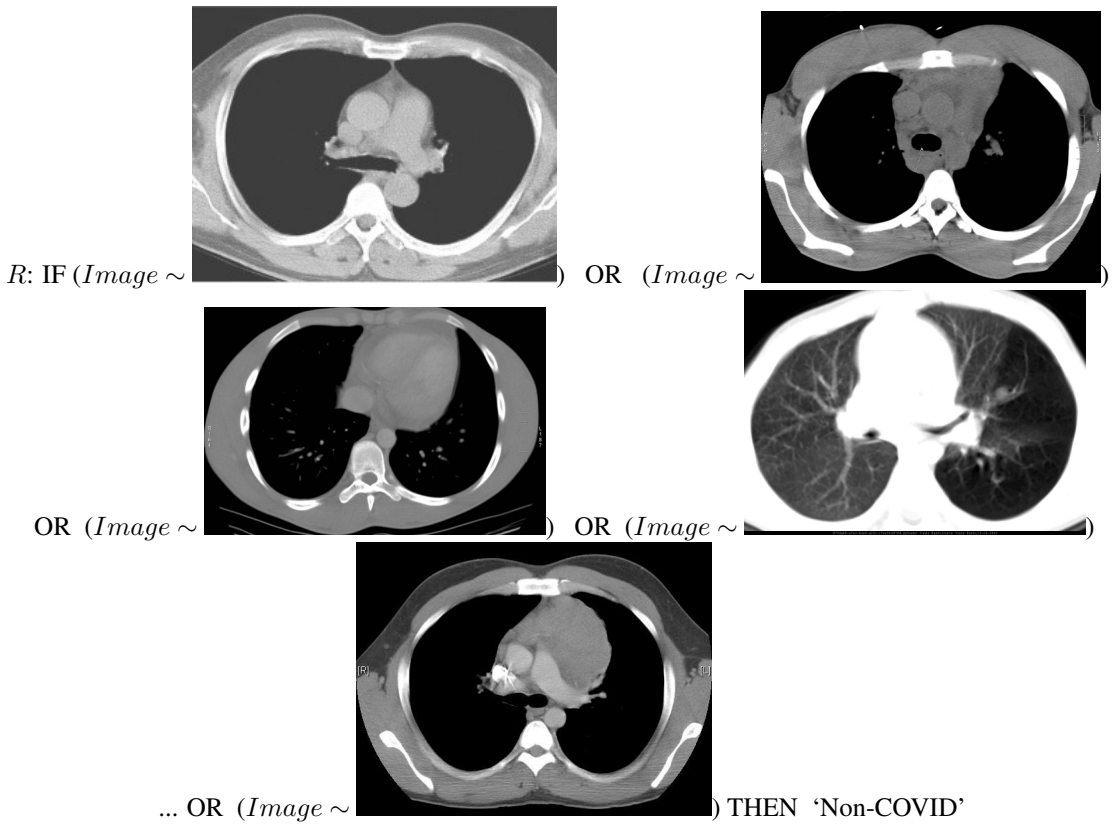


Figure 4: Non-Covid final rule given by the proposed eXplainable Deep Learning classifier.

retrospective cohort study. *The Lancet*, 2020.

[4] Catrin Sohrabi, Zaid Alsafi, Niamh O'Neill, Mehdi Khan, Ahmed Kerwan, Ahmed Al-Jabir, Christos Iosifidis, and Riaz Agha. World health organization declares global emergency: A review of the 2019 novel coronavirus (covid-19). *International Journal of Surgery*, 2020.

[5] Ensheng Dong, Hongru Du, and Lauren Gardner. An interactive web-based dashboard to track covid-19 in real time. *The Lancet infectious diseases*, 2020.

[6] Tao Ai, Zhenlu Yang, Hongyan Hou, Chenao Zhan, Chong Chen, Wenzhi Lv, Qian Tao, Ziyong Sun, and Liming Xia. Correlation of chest ct and rt-pcr testing in coronavirus disease 2019 (covid-19) in china: a report of 1014 cases. *Radiology*, page 200642, 2020.

[7] Yicheng Fang, Huangqi Zhang, Jicheng Xie, Minjie Lin, Lingjun Ying, Peipei Pang, and Wenbin Ji. Sensitivity of chest ct for covid-19: comparison to rt-pcr. *Radiology*, page 200432, 2020.

[8] Adam Bernheim, Xueyan Mei, Mingqian Huang, Yang Yang, Zahi A Fayad, Ning Zhang, Kaiyue Diao, Bin Lin, Xiqi Zhu, Kunwei Li, et al. Chest ct findings in coronavirus disease-19 (covid-19): relationship to duration of infection. *Radiology*, page 200463, 2020.

[9] Wei Zhao, Zheng Zhong, Xingzhi Xie, Qizhi Yu, and Jun Liu. Relation between chest ct findings and clinical conditions of coronavirus disease (covid-19) pneumonia: a multicenter study. *American Journal of Roentgenology*, pages 1–6, 2020.

[10] Weifang Kong and Prachi P Agarwal. Chest imaging appearance of covid-19 infection. *Radiology: Cardiothoracic Imaging*, 2(1):e200028, 2020.

[11] Ming-Yen Ng, Elaine YP Lee, Jin Yang, Fangfang Yang, Xia Li, Hongxia Wang, Macy Mei-sze Lui, Christine Shing-Yen Lo, Barry Leung, Pek-Lan Khong, et al. Imaging profile of the covid-19 infection: radiologic findings and literature review. *Radiology: Cardiothoracic Imaging*, 2(1):e200034, 2020.

[12] Heshui Shi, Xiaoyu Han, Nanchuan Jiang, Yukun Cao, Osamah Alwalid, Jin Gu, Yanqing Fan, and Chuansheng Zheng. Radiological findings from 81 patients with covid-19 pneumonia in wuhan, china: a descriptive study. *The Lancet Infectious Diseases*, 2020.

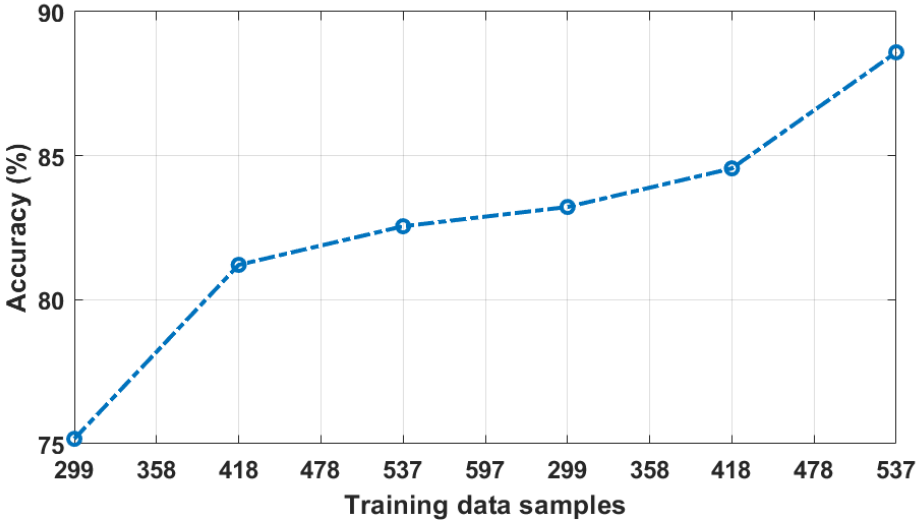


Figure 5: The figure illustrates the evolving nature of the proposed approach. It continuously learn as new training data arrives to the system. It can be observed that with 478 training data samples the proposed approach could obtain better results in terms of accuracy (84.56%) than the Baseline approach (84.0%) with 537 training data samples[19]. The Baseline approach is a Deep Neural Network which needs large number of training data to obtain high performance in terms of classification and once trained can not be further improved unless fully retrained. In contrast, the proposed approach can obtain higher performance using less training data due to its prototype-based nature.

[13] Daniele Ravì, Charence Wong, Fani Deligianni, Melissa Berthelot, Javier Andreu-Perez, Benny Lo, and Guang-Zhong Yang. Deep learning for health informatics. *IEEE journal of biomedical and health informatics*, 21(1):4–21, 2016.

[14] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[15] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[16] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5):206–215, 2019.

[17] Gary Marcus. Deep learning: A critical appraisal. *arXiv preprint arXiv:1801.00631*, 2018.

[18] Plamen P Angelov and Xiaowei Gu. *Empirical approach to machine learning*. Springer, 2019.

[19] Jinyu Zhao, Yichen Zhang, Xuehai He, and Pengtao Xie. Covid-ct-dataset: A ct scan dataset about covid-19. *arXiv preprint arXiv:2003.13865*, 2020.

[20] Qiang Du, Vance Faber, and Max Gunzburger. Centroidal voronoi tessellations: Applications and algorithms. *SIAM review*, 41(4):637–676, 1999.

[21] Feng Pan, Tianhe Ye, Peng Sun, Shan Gui, Bo Liang, Lingli Li, Dandan Zheng, Jiazheng Wang, Richard L Hesketh, Lian Yang, et al. Time course of lung changes on chest ct during recovery from 2019 novel coronavirus (covid-19) pneumonia. *Radiology*, page 200370, 2020.

[22] Yan Li and Liming Xia. Coronavirus disease 2019 (covid-19): Role of chest ct in diagnosis and management. *American Journal of Roentgenology*, pages 1–7, 2020.

Author contributions statement

P. A. conceived and detailed the idea. E. S. designed and implemented the algorithms, designed and performed the experiments. P. A. and E. S. wrote the manuscript and interpreted the results.