



Dr. Jong Chul YE

Use of AI for COVID-19 – Diagnostic and Patient Privacy

I'm Jong from KAIST and I'm from department of Bio and Brain Engineering. In fact, my background is computer science and working on biomedical engineering applications. So that's why I'm in the bio engineering, but I'm a computer scientist working on the AI. So I think, actually right now under the COVID-19 pandemic situation, many of the previous speakers have mentioned about the negative aspect of AI, but here, the topic is more like a positive aspect of AI under this kind of pandemic situations.

Especially I'm going to talk about the role of AI for the patient triage and the patient management perspective. And I will also talk about that, even though people think about AI can do anything, but under the current pandemic situation, learning the AI machine and training the AI machine has a technical difficulties and also some legal issues as well. So I will talk about these issues as well for the discussion.

In fact, as of today, when I look at the Internet, we already have a 9 million confirmed cases, nearly half million people died. In fact, I first prepared this presentation for another webinar a month ago; at the time, the case was half of this one. So within a month already we have a doubled number of confirmed cases and deaths. The reason for that is actually COVID-19 has a factor, so called the basically producing number or R_0 number. So for example, originally R_0 number means that one patient can infect three other people and the other three are the peoples can impact another nine people. Every time it tripled. So in fact, what I found is in all the studies in China during the January and February, what they found is that R_0 factor of COVID-19 is nearly four or even seven as you can see in this graph.

That means this is a real issue on transmission. And that's why now in the really short time period, COVID-19 become already global pandemic. The problem here is now we know as you know that we don't have any vaccines, or, we don't have any kinds of treatment. So because of that, the only way to deal with this COVID-19 pandemic and then flattening the curve is based on the early diagnosis and the

contact tracing. So one of the golden standard for the early diagnosis is so called PCR or polymerase chain reaction. This is based on the biochemical reactions. So the sensitivity is more than 90%, even 95%. But the main problem here is that it takes a quite long time, like six hours for the sample preparation and recharge, and the whole process can take like a half day. So we need to wait.

And furthermore, because of the global emergency situation, that there is a shortage of PCR. Because of that, in many of the countries, even though there are a lot of patients, it is still very difficult to get the PCR, and also, in the third world countries, the PCR is expensive. So because of that, using the PCR for all the early diagnosis can be a problem.

And in fact, another way in China to do the diagnosis is based on chest CT. The reason for that is during the early pandemics, the chest CT is very quick compared to the six-hour PCR. Likely, less than half hour, we can actually see whether this patient is affected by COVID-19 based on the images. Furthermore, the accuracy or the sensitivity of chest CT based, I would say, more than 90%. So this is a very sensitive system, but the problem here is, other than China, in most of the countries, the CT is very expensive to take. Furthermore, another more serious problem here is that, there is a potential for cross contamination. That means, if this is used for the diagnosis, not only for the patient's use, but also for other normal persons. In that case, there is a potential for the contamination from the COVID-19 patients to normal persons and healthy persons. So this is an issue.

In that sense, let's have a look at chest X Ray imaging. People are very interested in looking at this kind of thing, because it's very quick to take images. Furthermore, it's very cheap. And even in the third world countries, chest X rays have wider spread, but now the problem here is that the sensitivity of this one is not good. We've been with the experience, the rate is less than 70%, but one good thing people respond is even in the negative cases of PCR, there are some cases like about 9% of patients that had the chest X Ray, are positive. So that means they can actually detect the COVID-19 even in PCR is negative. So because of that, a lot of people are interested in using AI to improve this kind of accuracy of chest X Ray based diagnosis for the patient triage.

The reason for that is as you know, that COVID-19, it's a lung infection by the viruses. So because of that, you can see the CT images here, you can see a little bit of a gray area here. This is the early stage of infection of the lung infection. So that can be actually seen as a so called GGO in the X-ray images as well.

Furthermore, if we go, furthermore, in the serious cases, if you have a lot of serious infection in the lung area in the CT images that can be easily to see from the x-ray

images. So based on this kind of clue that this is a lung infection, people are interested in utilizing the X-ray or other imaging modality for the patient triage, etc.

What does it mean by the patient triage? Let's think about this kind of situation. For example, some people come with symptoms like fever and cough, will come to the emergency room, okay. Before they are admitted to the emergency room, you need to see that whether they have a COVID-19 or they have other types of virus. So one kind of way to think about it is, for example, you can actually do it directly using the RT-PCR, but the problem there is that, you need to wait for another half day. That means the patient cannot be admitted to the hospital. So one way to do is, taking a chest X Ray images and using the AI, for example, that they have like based on a bacterial pneumonia or tuberculosis, in that case, they can get the antibiotics from the doctors and can go home or some will be admitted to the hospital. But if the AI found that this is some viral pneumonia, in that case, they can use the RT-PCR or chest CT for another diagnosis. And if they are now diagnosed to have a COVID-19, they are now isolated and then admitted to a negative pressure rooms.

Now, by doing that one, you can actually save a lot of resources, for the use of this kind of PCR because of what I mentioned, and in this global emergency, this is actually very important and also saves resources.

Furthermore, another thing, this kind of triage it is possible, it turns out that even though a lot of people have a fever and non-infections, there are majority of people are coming from the bacterial pneumonia and tuberculosis. So if you have a very good AI machine to increase the sensitivity, to exclude those kinds of things. You can actually save the medical resources during this Pandemic.

And also not only these kinds of patient triage, you can also utilize AI for the patient management as well. For example, this is the study that has been done in China using the CT data, for example, using AI, from the CT images of the lung, you can identify the region of these kinds of infections, and then they can actually score it, what they call the Corona Score. And based on monitoring this kind of Corona score, they can actually see whether this patient is improving or need a more emergency cares and etc., they can manage those kinds of things. But without these kinds of objective quality by just looking at it, is a little bit difficult. So that's why AI can play the role in this case.

Another modality people nowadays are interested in, is ultrasound machine. So ultrasound machine is a handheld system. So they reach to you, and they actually read a response, some kinds of abnormalities in these images, because of lung infection. But the thing is compared to the CT and X-ray. It's not clear as you can see

here. So because of that, they're interested in utilizing AI to actually have an automatic diagnosis of this kind of thing. This is actually one of the directions in technical side to utilize AI.

However, even though AI has a lot of advantages to actually help patient triage and management, there are a lot of serious issues, especially during the pandemic.

The problem is AI or AI-related machine need a lot of data, but during the current global emergencies, it's very difficult to collect the well curated COVID-19 data set, because doctors are very busy in treating the patient. They don't have a time to actually curate the data set. Furthermore, depending on the region and depending on the countries, the age distribution of patients is different. For example, in the European country, most of the majority of patients are from the old generations and in Korea in the early outbreak of COVID-19, there were some religious factors, so because of that, the patients are more like 20s to 30s, that generation.

Recently, I heard that, nowadays in America, there is some concern about the patient population in children. So because of that, it is also need to have those data sets.

So that means this kinds of unbalanced age distribution, is actually a limited factor for a universal diagnosis for this kind of purpose.

So in fact, a group has been investigating this kind of thing from a technical side as well. In this paper, we actually demonstrate that using this kind of situation, using limited number of data, we can still improve it compared to the standard method. You can actually improve a nearly 20% of accuracy. And by doing that, for example, this is actually the AI-based diagnosis for the COVID-19, this is probability of COVID-19 infection, more red is more COVID-19-like problem, but compared to other bacteria or tuberculosis, or normal patient, AI can actually accurately localize the regions. And accurately diagnoses the probability of this kinds of COVID-19.

However, still, to have a good AI machine for this kinds of patient triage and handling, you still need a lot of data.

So because of that in the engineering communities, the globe community, there are a lot of papers are now collecting and releasing the public data for the research purposes. For example, Kaggle is a one of the important platform to show the public data, after removing the personal information of course, and recently the Spanish school released very huge database for the chest X-ray and CT data as well. And also, I know there are a lot of database from the Chinese side as well.

However, still with this kind of data, one of the issue is what the two previous speakers mentioned, the privacy issue. That's actually a very important issue in training the AI machine, because especially this is general data. So you need to collect data that actually can identify some of the patient data as well.

So to save those kinds of patient privacy in engineering community, in the AI community, there are a lot of techniques has been recently investigated. For example, one of the most actually important platform people are investigating for the healthcare application is so called, the Federated Learning.

In this case, this does not share data. So they just use their local data for the train, neuro network, and each hospital can train neural network using their own data, but they share only the way itself instead of data, because the way itself is a statistical summary. So because of that, in that case, you can actually save the patient privacy for that. So this is actually one of the important directions, especially to handle the medical data.

But still to deal with these kinds of COVID-19 pandemic issues especially, one of the issues here is that, data collection within a short time period is very important. So we need actually the global international collaboration, from the government side, as well, for example, WHO, and also patient and drug makers and radiologist, and physicians, and researchers like me, and etc..

In fact, for accelerating this research, we still need to think about the patient privacy as well. So one of the direction I am now thinking about is, combine these kinds of AI with the blockchain such that patient data sharing, they can save the privacy and also they can be rewarded using some kinds of cryptocurrencies as well, but still using their data for the AI training and to improve and to collect more database and then to improve the accuracy of the machines, so that you can actually have a more accurate diagnosis and patient management to handle this kind of global pandemic.

Thank you much for your attention.