Subtitles and transcriptions
Subtitles and transcriptions are available for selected materials for purpose of helping users understand the contents of the educational sessions.
Uncertain words have been indicated with ?? before and after the part. Parts that could not be understood at all have been indicated as [Audio Not Clear].

Every effort has been made to faithfully reproduce the audio of the sessions as recorded. However, no responsibility is accepted for mistakes or omissions. ESO does not endorse any opinions expressed in the presentations.

Artificial intelligence for prostate cancer radiotherapy

Prof Bibault: Hi everyone. Thank you for inviting me to discuss today about AI for prostate cancer radiotherapy. I'm Jean Emmanuel Bibault, and I'm a physician scientist, I'm working in Paris, and I'm also doing research in AI. So, what I want to talk to you about today, is actually quite old scientific field, which we call AI or artificial intelligence. The first concepts of AI and precisely neural networks, dates back to the end of the 40s. And you should know that it's actually a field, that has known some kind of golden ages, and also, dark ages that we also called AI winters, which were periods during which the scientific community was not really interested into AI. Because we used to think that AI was never going to be possible. We've seen, more recently, a new interest in AI and most importantly, in machine learning, and deep learning. Mostly because of the works of three people, one from Toronto, which is Geoffrey Hinton that you see here on your screen, on the lower right-side. And also, from Yoshua Bengio who's also Canadian from Montreal. And finally, from Yann LeCun who's a researcher from France, who is now the Director of AI at Facebook. So, these three persons were actually called the godfathers of deep learning, and they have really launched and renewed the interest in AI that we've seen from the beginning of the 21st century. Just to remind you that when we first started to do AI back in the 50s, we used to create networks of neurons who are very real. So, you have here on the right-side of your screen, an example of one of the first machines, which was called Perceptron, and this is one from 1951 or -2. But today, of course, all these wires, all these electrical wires have been virtualized. And now, you should know that AI is mostly done with frameworks, which are software frameworks, available from Google, such as Tensor Flow, or Facebook, such like Pytorch, or from computer sciences school, such as Scikit-learn, which is developed and released by a French well known school which is called INRIA. So, the goal today is to try to show you that AI is not some kind of fantasy that you read about in the journals. It's gonna be very, very real, and it's going to completely change the way that we do medicine, of course, and of course, within medicine, oncology. And this is just a few of the recent titles that you could have read mostly in the American journals, but I'm sure that in each of our countries, you've seen this kind of very catchy titles, and actually, they are not too far away from the truth. At least the truth that's going to emerge in the coming years. What I want to show you today is a few examples of works that we performed in the AI lab of the Stanford University, in the US. And that we have continued in our INSERM lab here in Paris. And these works are mostly about screening cancer, and in our case, prostate cancer. Diagnosis cancer and also planning treatments, evaluating, and assessing treatment response. And finally, you should know that AI can also be used for follow-up mostly through telemedicine. And you should know, for example, that just today, the ESMO has released the guidelines for telemedicine of patients treated for cancer. So, it's a very, very relevant subject. So, just a reminder, you can ask your questions in the Q&A button, and we will keep them at the end, and Pier Francesco and I will try to answer them as best as we can. So, the first part of our topic today is about prediction, diagnosis, and epidemiology. I want to talk to you about work that was performed in the US, I guess you know that each year in the US something around 200.000 new cases of

prostate cancer are diagnosed. And usually, what we do for that, I'm sure you know that, but we do PSA testing, prostate biopsy and then, we do staging in order to try to stratify the risk of these patients. Currently, our risk stratification strategies are quite primitive, I would say, because they depend on very simple variables such as PSA, the tumour stage, the lymph nodes, metastasis, or the tissue and organs of metastasis. And sometimes, in the best-case scenarios, or in the best centres, you can also use genomic signatures such as Decipher for example, when you decide to treat a patient with radiation therapy. All these data is used to classify patients within only four, I mean, global stratification, localised, locally-advanced, metastatic and finally, castration-resistant. So, you see that it's pretty basic, and we need to be able to go further than that to try to stratify the patients in order to better adapt treatments. Why is it important to adapt treatment and treatment strategies? I'm sure you've all read this great paper, in The New England Journal of Medicine, that was published a few years ago about the ProtecT trial, which was randomised control trial. And you see here, the results on the prostate cancer specific survival, where the authors compared three different strategies. The first one being surgery, second one radiotherapy, and the third and last one active monitoring. And you can see that whether you do surgery or radiation therapy, or even active monitoring, which is basically nothing, or almost nothing, prostate cancer specific survival is the same even when you have like a 10-year median survival, I mean, median follow-up. And it's really important to keep in mind that in that trial there were some very high-risk patients, including some T3 and T4 patients who were in the active monitoring arms. So, it shows that with our current strategies, we are not really able to discriminate the patients who are going to benefit from the treatment, from the patients who are not going to benefit from the treatment. Why is it very important specifically in prostate cancer to be able to know which patients are going to benefit from the treatments? Well, it's pretty basic, that's because our treatments, whichever they are, are actually extremely toxic on a urinary point of view, also on a sexual point of view, and of course, on a digestive point of view. And each of these treatments, I mean, surgery or radiation, have a very specific and different profile of toxicity. So, what we do currently, is that we almost treat everyone in the same way and we induce a lot of toxicity, including to patients who will never have any kind of benefit from these treatments. So, we need to be better at decision making in prostate cancer. Just an illustration of that is that among the 200.000 patients in the US, new patients every year, only around 26.000 will actually die from the disease. And it is estimated in different studies that up to 35.000 patients are actually being over-diagnosed and will go through unnecessary treatments that will cause complication, that will cause toxicity and these patients will not benefit from treatments. So, how can we use what we have, existing data to better stratify patients and to determine which patient will benefit from treatment and which patient will actually never benefit from that? Of course, we already have the answer to that. The answer you will say is nomograms. Nomograms, they already exist, there are lots of nomograms available and they usually predict progression-free survival or cancer-specific survival. Most of the time, most of the published nomograms, not everyone but most, rely on data from only one centre. So, they're not always well generalizable and most of those also use a regression model, which is in itself not really an issue, but I think that today we have better tools to better discriminate patients. That fact is not really the main limitation of nomograms, the main limitation of nomograms is that they do not take into account the comorbidities of the patients. So, whether you have like a very high-risk of cardiac death or cerebral stroke, that does not appear within the nomograms, right? So, if a patient is at a high-risk to die from a heart disease, but a low-risk to die from prostate cancer, this will not show in the nomograms. So, this is basically like you are trying to look inside a room through a very tiny key-hole, and we need to open the door to have a whole picture of the patients, and not only the prostate cancer variables. And this is an example of one of the most used nomograms, which is from the MSKCC New York, and if you go online and use that nomogram, they also warn you from these very facts that this nomogram does not exclude the possibility of death from other causes, such as heart disease or accident within the time-period that they were developed for. So, the main goal of the work that I want to present to you today was to address that challenge. To do that, thankfully, we have already a very important amount of data. And the data that we used was data that we extracted from the PLCO trial, that is a very well known, randomised multi-centre trial that was performed in the US, where approximately 80.000 men were randomised in 10 US

centres between annual screening for any kind of cancer within prostate, lung, ovarian and colorectal. And also, the control arm had usual care, so no screening. And you should know that within that trial, we have a very, very large amount of data, not only on screening, but on the whole outcome of the treatments of these patients on this very large number of patients. So, we had to draw data transfer agreement with the NCI, and we were able to access that data, which is a very granular and very high-quality data, because it's been gathered in a prospective manner, which is of course extremely important when you want to develop AI, because if you want to develop AI, you need very high-quality data, otherwise your model will be garbage basically. So, the dataset that was available from the PLCO trial is a very comprehensive dataset that contains nearly all the PLCO study data available for prostate cancer screening, incidence and mortality analysis. One record for each of the participants in the PLCO trial has baseline features, screening features, diagnosis features and all the treatment procedures. So, what we did and what we decided to focus on is, of course, patients that were actually diagnosed with prostate cancer during the follow-up of PLCO trial, irrespective of the arms that they were originally included into. So, the patient could be screened, and the patient could also be screened. There's been a bit of controversy about the PLCO trial, because we know that many patients from the control group who are not supposed to get any kind of screening actually got screening through digital examination or PSA. So, that's why we took patients from both arms to try to limit the biases from this kind of contamination. The next step that we performed was to assess and select the productive power of a very simple set of questions as a baseline indicator for prostate cancer specific and overall survival, 10 years after diagnosis. So, you can see here the whole list of the features that we extracted from the PLCO trial, they were from the prostate cancer diagnosis, very basic and [Audio Not Clear] features, but we also had the whole medical history of the patient, his physical activity, as you know, and there's been quite a few studies about that, physical activity is a very important predictor of toxicity, of course, but also of survival. Social economic status, which is of course also a very high predictor. And finally, hormonal status, which we also have within the PLCO trial. So, once again, if you have any questions, feel free to ask them using the Q&A button, and we'll keep that in mind for the end of the topic. Once we had selected these features, we trained a model, and to do that, of course, this is a little bit on the technical side of the subject. But we split the data within training and testing dataset because we need, of course, to train a model on data and then, test it on data that he's never seen, otherwise it would be contamination also and it would be a little bit like cheating, right? We treated this task as a classification task, so basically, we wanted the model to predict whether the patient was dead from any cause within 10 years of diagnosis of one of the models that we call the overall survival model. And we also trained another model, where we wanted the model to assess whether the patient was at risk to die from prostate cancer within 10 years after prostate cancer diagnosis. So, as you understood, we had tabular data, and we know that for tabular data, which is basically a table, right? The state of the art, the best algorithm, the best kind of AI that you can use is XGBoost, which is a gradient boosted decision tree. And I will not specifically enter within the technical consideration for XGBoost, but there is a huge literature about that, that you can read online. One of the advantages of using XGBoost is that this algorithm inherently handles the missing values. So, you don't have to treat or to replace any of the missing values to train the model. Which is a huge advantage of XGBoost, and which is one of the reasons why it's working so well. We trained the model on the training datasets and we tuned what we called the hyper-parameters. The hyper-parameters are a set of parameters within XGBoost that you need to tune, which means that you need to find the most optimal one, the best one. And we did that in a nested cross validated way, only within the training dataset. To do that and to do that in the most efficient way possible, we used a Biogen optimization approach, which is basically a way to test the best parameters from a probabilistic point of view, without having to test all the combination of all the parameters, which would take weeks and weeks. Another thing that we need to do when you use XGBoost, and when you do machine learning in general, that you need, and it's extremely important, to correct the class imbalance with positive class waiting. What is that? It's basically the fact that within the patients who were diagnosed with prostate cancer, there is only a small proportion of patients who will actually die. And if you have this kind of imbalance, an algorithm will, for example, risk to over-fit on this class. And so, you need to correct this when

you do the training in order to avoid that risk. Another thing that we wanted to do is that we wanted to provide, what we call interpretability. Why is it extremely important within medicine? It's important because most of the time and most of the criticism that machine-learning has is that AI is basically a black box, right? AI is extremely performant, it gives very good results, but you never know why the AI has given such or such results. No, well it's been a few years now, but we have methods to try to extract from the model, the reasons that the model has performed any kind of prediction. There are a lot of different methods and we chose to use one of the most effective one, which is called the Shapley values. And it's a unified approach to interpret tree models and XGBoost is of course a tree model. So, it it's perfect to use with XGBoost. What Shapley values do is that they reflect, and I will show you that later, the importance of every features on a population scale, so on the whole training dataset, but also, if you want at an individual scale. So, you can pick any individual and see why the model did this prediction, I will show you that. So, just to show you the advantages of Shapley values compared to other kinds of methods. After we did that, we deployed the model online, the two models. And we are able to provide very precise prediction, from a very simple set of features, between 20 and 30 features. And the patient or the physician can go online, enter the features and then, get his own prediction. So, what are the results of that approach? So, we used from the PLCO trial, all the patients who were the prostate cancer survivals, which is almost 9.000, which is a pretty good number of patients. And you can see here on that slide, the characteristics of these patients and what you should keep in mind is that most of the patients were localised or locally advanced. And we had actually very few metastatic prostate cancer that's basically because it was a screening trial. So, it's logical that we do not get a lot of metastatic prostate cancer. But you should keep that in mind when you want to use that algorithm. I think that it's not the best for metastatic patients. Otherwise, all the other features were pretty well balanced. And as you can see, a little bit over 500 patients died from prostate cancer and a little bit over 3000 patients died from any other cause. Here are the model's performances, and you can see that these performances are extremely high, and it's not usual to get this kind of very high performances. Usually, in publication, you've got very good AUC, which you see here, but very bad PRAUC, which is precision recall. And in our study, we had in both of these metrics, very good performances. And when you do an algorithm that tries to predict survival, it's extremely important to have both of these metrics, with a good value. So, it's another message for the presentation and for AI in general, do not only read about AUC. AUC is not very good metrics, most of the time. You need to consider all these other metrics in order to be able to interpret the performances of the models. I just wanted to show you both of these confusion matrixes. To show you that both the models are very, very good. And so, you want the models to have a very high number in that case, and in that case, in both of these tables, and you can see here that it's the case for both models. Another interesting thing is that we wanted to show that using this kind of model gives more interest to prostate cancer screening. I'm sure that you know that there is a lot of controversy regarding prostate cancer screening. Should we be screening patients for prostate cancer? Because we have data from the PLCO and from the European trial that are a little bit contradictory. The European trial showed that screening prostate cancer is good, I mean, from an overall survival point of view. But the PLCO trial showed that it doesn't change anything. So, what we wanted to do is take all the patients that were screened and not screened and apply the model to these patients. And we show that if you do this kind of two-step process, first, a conventional screening through digital examination or PSA and then you apply the model, we are able to discriminate with a very, very high difference between the patients who are going to be at very high-risk of prostate cancer death, and the patients who are not. So that can be used to better personalise treatment. And we showed even for overall survival, that these models combined with screening, can be very interesting. So, I talked to you about interpretability, and this is actually the result of the Shapley values on the populations scales. You have here, each of the features of the models that are taken into account into the model. And here, you have the contribution to the prediction. If the value is the lowest, it means that the risk of death is lowest. And if the value is highest, it means that the risk is higher. And you can see, for example, on the Gleason Score, that if the Gleason Score is very low, the risk to die from prostate cancer and even overall survival is quite low, but if the Gleason Score is higher, very logically, that risk increased. And with that, you are able to visualise that

your model is actually correct and compatible with the human intuition. So, you are able to actually check that the AI that you develop is not completely irrelevant and is not doing absolutely anything. And we show that on both of these models. And so, you can now go online and use these models online and get a very precise prediction on an individual scale that I'm going to show you now. So, just to check the consistency with our physician intuition, I wanted to show you one or two virtual patients. Let's talk about a patient who has a high-risk prostate cancer, but who has no significant comorbidities he's young, he doesn't smoke, doesn't drink alcohol, has a regular physical activity. And this is the exact results that you get online, you get the probability to die from prostate cancer, and you get also this little figure here that shows you in red the features that participate to increasing the risk to die. And in blue, the features that participate in decreasing that risk. In that case, in that patient, you can see that his risk to die from any cause including prostate cancer is around 20%. And his risk to die from prostate cancer specifically is also around 20%. That is easy to explain because basically this patient has no risk to die from any other thing than his high-risk prostate cancer. But if you take another kind of patient, with an intermediate-risk prostate cancer, and several comorbidities. He's older, he smokes, he drinks, he has no physical activity. And then, you will see that this kind of patient has a probability of dying from any cause of up to 25%. And you can see that's because he's a smoker, and he has no physical activity. But his probability to die from prostate cancer is actually extremely low. So, you need to understand that when you talk about risk of death, you actually talk about a competition between general cause of death and prostate cancer. What I mean is that if you are going to die from lung cancer, you will not die from prostate cancer, right? So, that's why it's extremely important to take into account the comorbidities of the patient. So, of course, these models are not perfect, but they were designed to answer, as I showed you, I think, a relevant clinical issue. Which is which patient will actually benefit from treatment and if so, which patient is going to die from prostate cancer. And if so, why? The why is extremely important. The results are pretty accurate and so, it's the first model that was developed using machine learning. And that was trained on such a large number of patients, from 10 different centres and from prospective data, which is extremely important. So, I think it's much generalizable, and by that, I mean it can be used on patients from other countries and other centres. Of course, the limits that I already mentioned, that's the trial, I mean the data that we used for the trial was not designed specifically for that. So, it can introduce possible biases. And most importantly, for patients with metastatic prostate cancer, there were only 2% of patients within the dataset. So, you should be very cautious when you use models for this kind of patients. Finally, of course there is a questionnaire response biases, because all the data within PLCO was based on questionnaire, right? And so, it could also introduce biases, but it's the same in any kind of study. So, the model has been deployed online. It's been presented at ASTRO in 2019, and at the ESMO in 2021 and it's being used worldwide. And if you want to use it, you have the URL right here on your screen. So, again, if you have any kind of questions, feel free to ask them. So, that was the first part, and the largest part about epidemiology, diagnosis, and prediction. One other work that I want to show you that is more related specifically to radiation oncology is treatment planning. And basically, you know that we can use AI, and most specifically, most frequently deep learning to do all these different steps which are segmentation of the tumour and the organs at risk, automatic dosimetry and, of course, ultimately, IGRT and adaptive radiotherapy. One of the works that we also performed in Stanford, is using deep learning and more specifically, sorry, U-Net network to automatically propagate the prostate segmentation from the treatment planning CT scan to the daily CBCT. So, as you know, when there is a patient who's treated with radiation therapy for prostate cancer, he usually goes through between 20 and 40 treatment sessions. And each day, or at least once a week, he's got CBCT, which is a volumetric imaging to try to check for his positioning and for his bladder rotation, for example. So, what we wanted to do is use the segmentation that we did manually on the treatment planning CT scan and try to use that to guide deep learning to automatically segment CBCT. And so, do that every day to better adapt the radiotherapy treatment. So, in order to do that, again, we used two groups, and for the time-sake, I will not go too much into details, but we used a group for training, a group for validation, and for testing. The cases were segmented by four different experts of prostate cancer. And then, we compared the performances of the unit compared to the performances of the humans. And we show with these metrics

that the unit is able to segment the CBCT with a very high-quality, compared to the observer. Of course, we could discuss, and I would agree, that the DICE score is not necessarily the best score, always the best score to assess the performances of an auto-segmentation algorithm, but still, it's a very interesting result. You can see here the example of three groups of units, and the difference between the human control, and the unit, the automatic propagated controls. Again, with the four examples of four observers and the consensus, the consensus is actually a way to average the controls of the four observers within only one. And so, I wanted to show you that because this method is actually very fast, very reliable and we could use that, it's still in the research stage, but ultimately, we could imagine that we could use that for daily IGRTR and daily segmentation, automatic segmentation of CBCT for prostate cancer radiotherapy. Okay so, what are the prospective in AI for prostate cancer? So, now I'm going to be a little bit more on the science fiction side. I guess that I've shown you a few examples, but there are many, many more of AI in diagnosis and radiation therapy treatment planning. But you should know that someday it's not completely unimaginable that AI will perform surgery. There are already works in Stanford, for example, to train a deep learning model to identify the surgical tools of a surgeon automatically on a video in real-time. So, if are already able to do that, we can imagine that in a few years, maybe 5 or maybe 10 years, the deep learning models will be able to see not only the surgical tools, but also the operating site. So, this is one of the examples that you can read in Nature Reviews Urology that discuss that example. But most importantly, beyond all that, we should not forget the limits of these methods. AI is not like a magical wand; you can read and you can create a very bad AI. And of course, you can also create very good algorithm. So, I think that we need to train the physician, from today and from the future, to be able to interpret the results of the published study and the results of the AI systems that are going to be available for us to use. We need validation, just like we need validation for drugs. I think that we need to perform dedicated trials with specific methods to be able to assess the actual performances in the real-world of any kind of AI system. Of course, it's extremely complicated to define which are the quality criteria of this kind of algorithm. And we are currently writing in collaboration with the AAPM, so the American Association of Medical Physics, the ESTRO-ACROP guidelines for developing AI within radiation oncology. So, I want to thank the team that welcomed me in Stanford and my team now in France. And of course, my department of Radiation Oncology here in Paris. And we can answer the questions now, I guess.

Prof Franco: Yes, sure, thank you very much, Jean Emmanuel for this great, great talk. And the topic is, of course, very intriguing and there's few questions actually coming from the audience. So, I would want you, eventually, to answer. The first one, I think, you mentioned when you were discussing about nomograms, that nomograms are mostly relying on regression model. And so, one of the attendees is asking what is actually a regression model, or maybe, I mean, how can you compare it, say, the methodology that you use when you regress on a linear regression or on a logistic regression, and what is the difference between the statistical approaches that you use with AI, for example?

Prof Bibault: So, regression models are basically some of the more simple models that you, most of the time in the studies that you can read online or in any kind of journal it's regression model. And they have the advantage to be also easily interpretable because what you get when you do a regression model is almost like an equation. And you see that each variable, each feature has upon direction, a weight that will influence the final outcome. This is not the case in machine learning, in statistical learning. You do not get this kind of simple, easily interpretable equation, and that's why you need to use external methods to assess the participation, the weight of each variable to the decision. So, if you look at any kind of nomogram papers, most of the time, not always, but most of the time, you can get the direct equation that you can use to input each of your variables and to get a prediction.

Prof Franco: Yeah, thank you. Thank you, Jean Emmanuel. The second question is about the difference between... if there's a difference, probably there are some of... kind of overlapping one into the other in terms of domain, the difference between AI deep learning, machine learning. These terms are somehow confused between each other. So, probably it'd be helpful to clarify.

Prof Bibault: Yes, absolutely. It's a very frequent question and a very relevant question because, well, to put it simply, AI is only a very large field, that means that you are trying to develop methods to emulate intelligence in a computer. Okay, so, I guess that everyone knew that, right? But within AI, there are already several categories, for example, I mentioned like the AI winters, and the AI that was developed in the 60s, 70s and at the beginning of the 80s. Most of the time, this kind of AI, including AI that were developed within healthcare, already at the time, are symbolic AI. What is symbolic AI? Symbolic AI is simply a set of manual rules that were defined by domain experts. So, who basically, you asked the physician, "please try to create rules that I can use to classify patients." And I will answer each one of these rules automatically, and this will give me a prediction or this will give me a classification. So, this did not rely on any kind of data. It only relied on the experiences and the experts, okay? There is another kind of AI which is called, which has many names, but which is basically called statistical learning or machine learning. And this is basically a completely different approach where instead of creating manual rules, you are going to give a large amount of data to an algorithm and hopefully try to train that algorithm to find by itself the rules that are within that data to classify the patient. So, you do not create any kind of rules, this is the algorithm that's going to learn by itself these rules, and these rules can be extremely complex. And so, they are most of the time not easily interpretable. Within statistical learning, there is lots of different methods. For example, there is XGBoost, that I mentioned, which is a gradient boosted decision tree. A decision tree is basically a set of rules and the algorithm is going to define each one of these rules by itself, and it's going to give you like a prediction by answering each one of these automatically defined rules. But there are many other methods, for example, there is support vector machine, where the algorithm is going to try to find the most representative points, data points, and then, classify each of the other points depending on these two support vectors. And there are also KNN et cetera, there are a lot of different methods. And then, within machine learning still, there is what you call neural network. A neural network is another kind of machine learning, another kind of statistical learning, where we tried to emulate the workings of a human brain, I mean, of a brain. And so, you are going to define units which are neurons, which are connected between each other by, like that's where I showed at the beginning in the perceptron. It used to be wires, but now it's software. And the neurons are going to define automatically their own weights. And it's going to define their participation into the final prediction. And then, within neural network, you've got deep neural network. And so, it only means that you have a very high number of layers of neurons. And that's why you say that it's deep, because the deeper you go within the layer and the bigger the network is. So, I guess, I hope I'm clear, but it's actually pretty simple. Deep learning is a category of neural network, which is itself a category of machine learning, which itself is a category of artificial intelligence.

Prof Franco: Yeah, they are somehow comprised within each other, right?

Prof Bibault: Yes, a little bit like Russian dolls basically.

Prof Franco: Correct.

Prof Bibault: And just to close on that, it's pretty interesting because we've been, lately, concentrated, focused a lot on deep learning. If you like go on Facebook or your iPhone or whatever, there is a lot of deep learning going on. And we are suspecting that we might be hitting a wall, I mean like, we are not necessarily able to go any further within the performances of deep learning. And so, some researchers are considering that we should be trying to mix a symbolistic approach, with a deep learning approach. And so, this is going to be the development, I guess, in the coming years is to try to mix human expertise and machine expertise.

Prof Franco: Good, good. Thank you, Jean Emmanuel. One quick question is how is the DICE index, how is it calculated?

Prof Bibault: Yeah, the DICE, it's a pretty basic index. I mean, it's like a similarity index. Where you compare the control of… the automatic control that was obtained from deep learning, and the control that was

obtained from the human operator. And then, you are going to compare the union and the addition and then, you divide each one of these and you get a DICE. And the higher the DICE is the better it is.

Prof Franco: Yes. So, it's basically intersection divided by the union.

Prof Bibault: Exactly.

Prof Franco: So, if the volumes are coincident, then it's one, and actually, it's multiplied by two.

Prof Bibault: By two, yeah. But once again, DICE is not necessarily the best index. There are many more indexes and there are actually quite a lot of publications talking the best index that we should use but still, most of the publications report to DICE index.

Prof Franco: Yeah, DICE index is highly volume dependent. So, it really depends, the performance really depends on the volume of the structure that you're comparing. And there's another question that just arrived. Does deep learning include also radiomics and genomics?

Prof Bibault: Deep learning is only a method to analyse a set of data, to create a model. Radiomics is a method to extract quantitative information from a morphological image, and genomics is a way to extract data from sequencing, genomic sequencing. So, deep learning can be used to analyse radiomics or genomics, but it's not on the same level.

Prof Franco: Yeah. It's something like an instrumental method to work on the others. Great, great. I have a couple of questions, just quickly for you. I think you made some like important statements, right? One was about the need for high-quality data for artificial intelligent processes. And this is like kind of garbage-in, garbage-out, kind of, let's say, approach, because I think there's some sort of misperception, right? Because whenever, for example, we talk about real-world data and observational data. That sometimes are not good quality data. And I think sometimes the researcher has the feeling that, okay, I have this kind of noisy data, not very high-quality. I can just use some like sophisticated neural network, deep learning technique, and I will be able to find like meaningful data out of this like trashy type of... meaningful results out of this trashy type of data. But you stated that it is not the case, right?

Prof Bibault: Yeah, well, not everybody agrees on that, but I think that if you have bad data, you can do whatever you want. It's just not going to work and it's gonna include a lot of biases. So, I guess, you should be, of course, always careful when interpreting this kind of studies. Including the one I showed you, we need to be very modest with the results that we get.

Prof Franco: Right, so good attention on the way you collect data and the data you're managing.

Prof Bibault: What is important is the amount and the quality of data.

Prof Franco: Right, right. And the advantage of managing data with artificial intelligence, you think is only a computational advantage in terms of like managing a big amount of data or the quality of the prediction, the quality of the result is also dependent on other characteristics of AI?

Prof Bibault: Absolutely, the quality of the prediction is highly reliable on the type of model that you use. As I said, for example, XGBoost is very, very good. But if you, we know that XGBoost is better than SVM or KNN in many situations. And we also know that for tabular data, for table, deep learning it's not like the best methods. Deep learning is mostly for image analysis tasks probably. So, yes, you need to adapt the kind of algorithm that you use. And again, it's not like a magical wand.

Prof Franco: Yeah, you need expertise and you to be used to work on that. So, sure. And the last question I think, because I think we're running out of time. You nicely show how the model you use were kind of pretty efficiently able to discriminate the type of patient who has like an excess in the risk of dying of prostate cancer compared to others that have high-risk of dying of competing causes of death, of comorbidities. And

you showed like an example of an, let's say, high-risk prostate cancer in a patient with, let's say, good performance status, or without other competing causes of death. And then, you showed let's say a low-risk or intermediate- risk patient with like heavy comorbidities that might increase the likelihood for him to die of other causes. So, this is quiet, let's say, I would say, easy scenario to discriminate with. But I was wondering what is the performance of the model for, let's say, low-risk cancer patient in a good performance status patient, with a long expectancy, so, where you potentially might go to surveillance, but you might end-up in having late relapse, that you are not able to predict. And the other I think challenging and tricky clinical scenario is having a high-risk prostate cancer in a highly comorbid patient. So, there, how the model is able to discriminate this situation and to predict the best strategy?

Prof Bibault: Yes, absolutely, absolutely. That's a very valid point. The performances are still pretty good. And once again, the advantage of these models is that they're going to show you the features of why the patient is at risk to die or not. So, you can use these features, if you want to try to tailor your decision. But before you do that, of course, you would need like a very rigorous and thorough validation of this kind of model in a more prospective matter. And this question actually raises the question of what do you do with information about the future? I guess everyone has seen movies such as "Minority Report". What do you do when you think that a patient is going to die from this or that cause, but if you actually adapt your strategy to that, how are you going to know that the algorithm was right or wrong and that your decision actually had any participation into changing that? And this kind of question, we clearly do not have the answers to. And I think that it's going to be a major, major issue in the use of mostly predictive AI. Which is a lot of the type of use that we are projecting... AI is going to be used for, I mean. And I don't have the answer to that, but I think it's really very interesting.

Prof Franco: Great, great, great. That's a nice, nice discussion, very nice elements for us think about. So, I think, Jean Emmanuel, I think for the sake of time, we need to close the session if there's no other urgent question from the audience, and I would want to thank you very much for this very, very nice talk and interesting topic.

Prof Bibault: Thank you.

Prof Franco: And I would want to thank all the attendees for joining us tonight. And e-ESO, of course, as always, thank you very much.

Prof Bibault: Thank you, bye-bye.

Prof Franco: Bye.