

第3回AIを活用したプログラム医療機器に関する専門部会

日時 令和4年12月5日(月)
15:00～
開催形式 Web会議

<開会>

- 事務局（浜岡先端技術評価業務調整役） それでは第3回AIを活用したプログラム医療機器に関する専門部会を開催いたします。本日はお忙しい中お集まりいただきありがとうございます。

<委員出席状況報告及び配付資料確認等>

- 事務局（浜岡先端技術評価業務調整役） まず委員の出席状況について御報告いたします。当委員会17名の委員のうち、現在13名に御出席いただいております。全委員の過半数に達していますので、専門部会規程第7条に基づき本委員会の成立を御報告いたします。欠席の御連絡をいただいている委員はおりませんが、陣崎委員、武田委員、森委員は遅れておられるとのことです。

次に配付資料の確認をいたします。議事次第、資料目録、資料取扱区分表、資料1～3、参考資料がございます。事前にメールで送付しておりますが、不足等がありましたら事務局まで御連絡をお願いいたします。

次に資料取扱区分表を御覧ください。資料は内容に応じて取扱として「厳重管理」「取扱注意」「その他」に分類し、それに応じた対応を取ることとしております。本日の配付資料1と参考資料1についてはその他に該当しますので、委員各自で適切に保管・管理・廃棄をお願いいたします。資料2、3については取扱注意のため厳重に保管し、コピー等の複製、第三者への開示は御遠慮くださるようお願いいたします。

それでは佐久間部会長、議事の進行をお願いいたします。

<AIを活用したプログラム医療機器に関するご講演と意見交換

「医療情報データベースの構築における法的課題」（殿村委員）>

- 佐久間部会長 皆様、お忙しい中お集まりいただきありがとうございます。よろしく御願いいたします。それでは議題に移ります。10月24日に開催された第2回専門部会ワーキンググループの参加者と概要を資料1にまとめております。論点については本日後半で議論しますが、まず最初に殿村委員から御講演をいただきます。第2回専門部会でデータベース作成と個人情報保護法について議論がありました。そこで弁護士の殿村委員にその観点での留意点等の御講演をいただきます。殿村先生、よろしく御願いいたします。

- 殿村委員 議論があった中で法的整理が必要なことがあったとのことですが、私のほうでその問題意識を全て把握できているか分かりませんので、仮定の

事例を置いた上で法律の適用関係がどうなるのかといったところを少し整理し、その後10分間のディスカッションのお時間をいただいておりますので皆様が現場で悩まれている事項等がございましたら御指摘いただければそこでディスカッションさせていただければと思っております。本日基本的には、医療情報に基づいたデータベースを構築する際の個人情報保護法上の論点、検討事項について御説明いたします。

まず患者等からインフォームド・コンセント等を通じて医療機関が個人情報、医療情報を取得する、その医療情報、医療データに基づいてデータベースを構築するというケースを考えてみたいと思います。理屈の上ではいろいろなパターンが考えられます。例えば医療機関が自らデータベースを作成することも理屈の上では考えられると思いますし、その医療機関自身がデータベースを使う、若しくはその作成したデータベースを製薬企業等にデータベース提供契約に従って提供するケースもあるかと思えます。

実務上は医療機関自体がデータベースを構築することが諸般の事情で難しく、例えばデータベースを構築する事業者はその加工を依頼する、すなわちデータベースの作成を依頼・委託してそこで出来上がってきたデータベースを医療機関自ら活用する、若しくはこのデータベースを製薬企業等に提供するケースもあろうかと思えます。若しくはこのデータベース事業者が主体となり複数の医療機関からデータ提供契約に従ってデータを収集し、自らデータベースを構築してそれを製薬企業等に使用させる、販売するといった形態もあろうかと思えます。

いずれの場合であっても各当事者は、個人情報保護法、あとは倫理指針の適用がある場合は当然これらを遵守する必要がありますし、忘れてはいけないのがここに書いておりますとおり各当事者間で契約関係があるはずなのでこの契約上の制約事項にも服する必要があるところが3つのチェックポイントになります。

個人情報保護法のおさらい的なことなのでサラッといきたいと思えます。個人情報の定義は、資料記載のとおりですが、容易照合性が認められる場合も個人情報として扱われる点は注意が必要です。病歴やお医者様等による検査結果等が要配慮個人情報というカテゴリに当たります。そういった個人情報がコンピュータ等を用いて容易に検索可能な形で体系的に整理されている場合にはそれらが個人情報保護法上の定義語ですが「個人情報データベース等」という形になり、そこに含まれている個々の個人情報のことを「個人データ」と呼んでいます。ここで注意が必要な

のは、「個人情報」という定義語と「個人データ」という定義語は別のものとして扱われていることです。個人データは、あくまで体系的に整理されていて容易に検索可能な状態になっている個人情報のことを指します。

個人情報保護法は、御案内のとおり個人情報取扱事業者に当たる人に適用があるものになります。ただ1点留意が必要なのは、いわゆる国公立の病院や大学等については従来は個人情報保護法ではなく行政機関等に個別に適用がある法律が適用されていましたが、今年の4月1日からこの法律が統合され、基本的には国公立の病院等にも同じ個人情報保護法が適用されることになります。ただ更に注意が必要なのは一定の事項、特にここでハイライトしております行政機関等匿名加工情報の提供等については、公的部門には一般企業のルールとは違うルールが適用されます。具体的には、募集・提案等のプロセスを経た上で活用する必要があるというルールが適用されるので注意が必要です。

データベースを構築する上で個人情報保護法上留意が必要な規制が幾つかあります。まずそもそも個人情報を取得する際に原則として取得に関しては個人の同意は不要ですが、要配慮個人情報を取得する際には同意が必要です。また利用目的の通知等が必要になります。注意が必要なのは個人情報の利用目的です。利用目的を特定しなければいけないという規制になっています。そしてその特定された利用目的の達成に必要な範囲でしか使ってはいけないですし、その利用目的を後で変更するのであれば当初の利用目的と関連性を有すると合理的に認められる範囲でなければなりません。

したがって当初健康情報等を取得する際にデータベースを構築するという利用目的を特定していなかった場合には、多くの場合はこの関連性を有すると言えないと思いますので利用目的の変更もできない。そうすると、そもそもデータベースを構築するというのは当初の利用目的に入っていない可能性が高いということになります。その場合は幾つかの例外を主張することを検討する必要があります。典型的なものとしては、いわゆる学術研究の例外。2つ目は今年4月から始まった匿名加工情報に加工する。3つ目がいわゆる従来からある匿名加工情報に加工するという方法です。これらを用いることで、当初の目的に入っていない二次利用も可能になることがあります。

もう1つは、第三者に提供する際の規制です。こちらも御案内のとおり原則本人の同意が必要となっております。例外は、国内に対する提供

と海外の第三者に対する提供で異なっています。国内に対する第三者提供については、1つ目が委託や共同利用という形での提供、若しくはオプトアウトです。ただこのオプトアウトに留意が必要なのは、要配慮個人情報対象外で、医療情報の多く、少なくとも医療機関等が持っているものはほとんど要配慮個人情報に該当する人が多いのではないかと思いますので対象外となり、個人情報保護法上ではなく次世代医療基盤法という法律に従いオプトアウトの形で出すことが可能になります。また、学術研究と匿名加工情報、これは先ほど利用目的のところでも出てきた例外です。

海外について、EUとイギリスについては、日本と同等の水準の個人情報保護法制を有している国として指定しておりますので同意がなくても出せます。それ以外の国についても、ここに基準適合体制と書いておりますが、契約をきちんと締結し提供先に一定の義務を課すことで同意なく出せるとなっております。あと学術研究の例外や匿名加工情報の例外もあるのは、先ほどの国内の場合と同様です。

先ほど仮名加工情報と匿名加工情報のお話をしましたが、この仮名加工情報というのは今年の4月1日から新しく始まった制度ですので、念のために御説明いたしますが、従来あった匿名加工情報よりも加工基準が緩やかです。左側が仮名加工情報ですがこの3つの処理を行うことにより、仮名加工情報に充てることができます。仮名加工情報と匿名加工情報を図示してみましたが、ある種、匿名加工情報のほうが加工程度が大きく、もはや個人情報ではないこととなります。他方で仮名加工情報は、加工の程度に応じて個人情報でもあり仮名加工情報でもあるというものと、もはや個人情報でもなくなっている仮名加工情報というものがあります。いずれにしても、加工程度が匿名加工情報のほうが大きな加工を加えていると思っていただければと思います。

仮名加工情報にするメリットの1つとして、先ほどのスライドで御説明したとおり、まず利用目的の制限を受けなくなる。より正確に申し上げると、利用目的の変更が自由にできるようになる、関連性を有していなくても変更ができるようになることが1つ大きいです。ただ1点留意が必要なのは、この仮名加工情報という制度は原則第三者提供ができないこととなっております。すなわち自社内、自機関内で活用するための制度として作られたものになります。更にひねりが入っており、第三者提供はできませんが委託、共同利用という例外は使えるということになっています。例えば仮名加工情報に加工した上で共同利用をすることにより、

事実上第三者にこの仮名加工情報を使わせることができるというのが現在の個人情報保護法の規律になっております。時間が経過しそうですので、この委託の限界、共同利用の限界の話は、後ほどのディスカッションでこちらに議論が及ぶようであればもう一度触れることにして、飛ばしたいと思います。

本日は駆け足で御説明いたしました。医療データからデータベースを作る際の個人情報保護法から見た検討のポイントとしては、まずそもそも対象となる個人情報に適法・適切に取得されたものかどうか、そしてデータ取得時の利用目的にデータベースの構築が含まれているか、恐らく多くの場合は含まれていないのではないかと思いますし、利用目的の変更というのも合理的な範囲内では難しい場合が多いと思いますので、例外を検討する必要があります。

学術研究の例外が1つありますが、その名のとおり学術研究目的でしか使えませんので商業利用のためには使えません。そうすると仮名加工情報、若しくは匿名加工情報に加工することで利用目的の制限を外すことが考えられますが、先ほど申し上げましたとおり匿名加工情報のほうが加工程度は大きいのでどの程度の加工の粒度であれば沿っているかを検討する必要があります。

また第三者提供の規制との関係では、自ら個人情報を持っている人が構築するのか、若しくは第三者に構築してもらうのかによって違います。第三者に構築してもらう場合は必然的に個人データの第三者提供の規制が必要になり、原則同意が必要ですが多くの場合は事前に同意を得ていないケースがまだ多いですので、例外を検討する必要があります。

こういった形で個人情報保護法というある種一般的に個人情報を扱っている事業者にすべからず適用される個人情報保護法上の例外を駆使すれば、医療データについても第三者に提供した上で一定の加工を加えてデータベースを構築すること自体は可能にはなっております。ただ一般の個人情報と医療情報を本当に同列で考えてよいのかというところは常に議論の対象になっております。

特に今年の4月から制度が始まった仮名加工情報に関しては、先ほど申し上げましたとおり、仮名加工情報に加工することで利用目的の制限を受けなくなるというメリットがありますし、第三者提供はできないのですが共同利用という形で事実上第三者に渡すことができるということで、非常に便利な制度であると捉えられておりますが、これを医療情報にもそのまま適用してよいのかというところは、正に今、厚生労働省でも議

論が進んでいるところです。この個人情報保護法という枠の中だけで考えるのではなく、倫理指針との関係、そもそも医療情報の特殊性を踏まえた議論が併行してなされており、そちらにも留意した上でデータベースの構築をしていく必要があるかと思います。一旦、私からの説明は以上です。

○中岡副部長 御発表ありがとうございました。いつも議論になることだと思うのですが、仮名加工と匿名加工の違いというのはきちんと整理されているのですか。

○殿村委員 仮名加工と匿名加工の違いは、どこまで加工すればこれに当たるのかという加工基準の点と、加工をして出来上がったものがどういうメリットを受けるのかという点です。まずどういうメリットを受けるのかという点は、法律上明確に書いてあります。匿名加工情報というのは従来からあるほうですがもう個人情報ではないという整理です。ですので利用目的の制限も受けませんし、第三者への提供も同意なくできるということの効果があります。

一方仮名加工情報も、匿名加工情報ほど加工しないのだけれども一定の加工を加えるというものです。効果としては、まず利用目的の変更が自由にできます。それと、第三者提供はできないのだけれども、委託や共同利用という例外は使えるという効果自体は明確になっています。

よく議論されるのは、加工基準がよく分からないということです。どこまで加工すれば個人情報保護法上の匿名加工情報に当たるのかというところは、一応法律上は要件は書いてあるのですが、一番下の「その他の適切な措置」という所にキャッチオール条項のようなものもあることもあって、対象の情報によってどこまで加工すればこれに当たるのかが分かりにくいという批判が従来からありました。

一方でこの新しい仮名加工情報のほうはかなりクリアではありまして、氏名等の特定の個人を識別する記述と、個人識別符号と呼ばれている指紋とかDNAといったものを消してしまえば財産的被害が生じるおそれというのはクレジット番号などがもう一つ想定されているのですが、医療情報では関係ないので基本的には非常に加工基準としてはシンプルです。ですので匿名加工情報よりも仮名加工情報のほうが、加工基準としてはシンプルだという違いがあると思います。

ただ「仮名化」とか「匿名化」という用語が倫理指針で使われていたこともあり、両者、若しくは仮名化した情報と仮名加工情報は法律上は違うのかといった用語の混乱から生じる問題というのは従来から指摘され

ていたとおりでと思います。

○中岡副部長 我々が議論すべきは医療情報で、画像やカルテということですよ。先ほどの話を聞いていると仮名加工情報はとても良いように聞こえるのですが、具体的にそういうものを上げた場合、どこまでやれば仮名でどこまでやれば匿名なのかという具体例がないと、そのレベルによっては結局匿名とほぼ同レベルになってしまう気もして、その線引きがはっきりしないとこれは本当は仮名レベルなのだけれども、匿名にしたから第三者利用してもよいよねと思われる方が出るような気がします。そこら辺について具体例を作られる予定はないのでしょうか。

○殿村委員 私が認識している限りは、今現在で公のものとして個人情報保護委員会などが出しているものはなかったと思います。もしかしたら個別の業界団体が出しているものはあるかもしれないのですが、そういったものが必要だという議論は幾つかの団体でも行われていて、私のほうでも聞いていますので、作っている、若しくは作ろうとしているという動きは一部にあるとは認識しております。

○中岡副部長 ありがとうございます。

○中田(典生)委員 1点確認なのですが、御存じのとおり、最近は画像合成が非常にできるようになったので、合成した画像は法律上どのような扱いになるか、個人情報ではないのか、それからまだ法律が及ばない点なのかどうかというのが1つです。

それから画像に及ばず電子カルテの情報で、患者の病歴とか臨床検査データまでを含めて合成するようなものを作るという研究もそろそろ出だしているの、そういうものも含めて日本の法律ではどのような範疇に入ればよいのかということをお教えいただければ幸いです。

○殿村委員 知識が及んでいなくて「合成したもの」というのが具体的にどういうものなのか分からないのですが、1点お伝えできることがあるとすると、その画像自体から特定の個人を識別できるかどうかという問題とは別の問題として、画像とID、この画像の人は01というIDの人で、それとは別に01というIDの人は殿村だという情報が仮に両方揃ってれば、この画像の人物は殿村ということは特定できることになりますので、問答無用にこの画像も個人情報に当たります。ここの定義で言うと②に当たり容易照合性と言うのですが、他の情報と容易に照合することができてその結果特定の個人を識別することができるようになる情報、すなわち今の例で申し上げますと、他の情報というのはID01は殿村であるという情報で、これとIDを通じて容易に照合した結果個人を識別することができること

となる情報というのは、今の画像情報になります。病院の中では多くの場合でつながっていると思いますので、単体で見たときは一見個人が識別できなくてもカルテ等の情報とひも付いていれば個人情報になるということは、1点お伝えできることです。組み合わせている、合成している場合というのが、どういうケースなのかは分からなかったです

○中田（典生）委員 イメージしていただければと思うのですが、例えばゴッホ風の絵を描く AI がありますがそれはゴッホ風だけれどもゴッホが描いたわけではないというわけです。それと同じように医用画像もその病気のものを全部合わせてシャッフルして部分的な特徴だけを出して絵にすることができるとい技術です。

○殿村委員 そうすると出来上がったもの自体は、本当に特殊な病気でない限りは誰の情報をベースにしたのかというのは出来上がった情報からは分からずかつ、先ほど私が申しましたひも付けがないのであれば一見すると誰か人の画像ではあるように見えるけれども実在しない人の画像であるということで、個人情報ではないと言える場合があるかと思えます。

○佐久間部会長 例えば MRI の画像と CT の画像があってそこから超音波の画像を生成してしまうというイメージを抱いたときに患者のデータが入力となって、そこで新しく関連したものが出てくるというイメージですね。

○中田（典生）委員 そうですね。既に、CT の画像から MRI の T1 強調画像を構成する技術は臨床で実際使われています。そういう合成技術というのは医療機器として使われ始めているので、そういうことも含めた議論が必要かと思えます。

○殿村委員 まず出来上がった情報が定義で言うところの①の個人情報に当たるのかという判断をして、今の例ですと超音波であるとか加工して特定の人のものではないということになるとまず①には当たりません。

次に②になりますが、元になっている個人情報とのひも付けがある種残っている場合には、例えば A さん、B さん、C さんの画像を合成したものだという情報が手元に残っている、若しくは出来上がったデータ自体からそれがある程度読み取れてしまうという場合には②を満たして個人情報に当たる場合もあります。

事例は違うかもしれませんが、例えば個人情報の A、B、C という情報をベースに AI を作り出すときに、A、B、C という個人情報から新しいものが出来上がっているのですが、AI プログラムとかパラメータと呼ばれているものは記号だったりするのでそれ自体は個人情報ではありません。個人情報由来で出来上がるものなのだけれども個人情報ではないという

整理になっているので、今のものもそのように整理できるものもあるかなと思いますし、この定義を満たしてしまうと個人情報に当たるものもあるのかなと思います。玉虫色な話ですが。

○中田(典生)委員 ありがとうございます。

○村垣委員 中岡先生の疑問である匿名加工情報と仮名加工情報の違いについて、非常によかったと思うことは、医療機関では合わせ技で個人情報と言われていたこと、例えば手術画像に全部名前は伏せていても手術日が入っていると何々病院で何日に手術をしたということでも個人の特定ができるだろうということで、個人情報とされていたのです。それが、仮名加工情報だと、手術記録で個人の名前などを外していればOKで、日付から特定できるから（あわせた情報で）不可となるのではないかというのが、具体的な例かなと思いました。次に、仮名加工情報の中で、個人情報であるというものと個人情報でないものがあるので、区別を教えてほしいということが私の質問の1点目です。

もう1つは個人情報とカルテ情報が結構混同されているところがあって、先日メディアを賑わせた手術ビデオを出した件とか確かに患者の臓器情報はビデオに入っていて問題となりましたが、カルテの自由記載の2号用紙の文言とか、医師あるいは医療従事者が行っている行為自体の取扱いはどうなるのか、行為に関するデータベース等の取扱いもありますので、患者情報とともに、医療従事者が行っている行為自体の情報に関しては、どう取り扱っていけばよいのか。この2点はいかがでしょうか。

○殿村委員 先ほど私がこの図で、個人情報である仮名加工情報と個人情報でない仮名加工情報があるとお伝えしたことに関連してだと思のですが、少し事例を変えてしまいますが、ID01の人は殿村であるという情報が1つあり、ID01の人が何月何日にこういう手術をしたという情報がまた別にあったとします。そしてこの手術をした人は殿村であるという、この「殿村」という名前を消したとします。そうするとこちら側の情報としては、ID01の人が何月何日にこういう手術をしましたというだけです。例えばその日に同種の手術が幾つか行われているのであれば、その情報だけを見ても特定の個人は識別できない状態になっています。そうするとこの情報自体は仮名加工情報に当たります。ただID01を斜交いにして、私の手元にはID01は殿村だというデータをもう片方の手に持っていますので、これを照合するとこの手術を受けた01は殿村だということが結局分かる。そうすると先ほど仮名加工情報になると申し上げたものは、個人情報でもあるということなのです。

更に、手術をした人が殿村であるということと ID01 であるということも全部消しますと、まず仮名加工情報になります。かつ ID も消してしまっているので何月何日に手術した人は誰かというのが他の情報とどう照合しても分からないというところにまでなると、これは個人情報でない仮名加工情報になります。

○村垣委員 再現できるかどうかですね。

○殿村委員 そうです。他の情報と照合して個人情報化できるかできないかという違いです。

○村垣委員 もう1点、先ほどの行為自体とか、いわゆる記載したものに、著作権があるのかどうかという課題もあるかと思うのですが、それと個人情報の関わりはいかがでしょうか。

○殿村委員 私の理解では、手術画像の問題というのは行為そのもののという捉え方と言うよりも、その動画だったり切り抜いた画像というデータ、そこはある種、人の情報なわけです。人に対する手術をしている情報ということで、もしかするとその情報だけだと術野しか見えていないので個人は識別できないかもしれないのですけれども、結局手術をしている病院からするとそれは誰かというのはカルテなど何かの情報とひも付けた上で手術をしているはずなので、どこの誰かが分からない形で手術をしていることはあり得ないので、そうすると提供した当該データだけを見たときには個人情報ではないように見えても提供元の病院からすると他のカルテなどの情報と照合して、この手術をしている動画の人は何さんだということが分かった上で提供している場合には提供した情報だけを見てそこに個人情報が入ってなくても法律上は個人情報を提供しているとみなされてしまうのです。これは提供元基準と言うのですが、提供元では個人情報でこのデータだけが提供先に行ったときには、提供先では照合はできないので個人情報ではなくなるという場合にそれを個人情報の第三者提供と見るのか見ないのかというのは、提供元を基準にするのか提供先を基準にするのかで違うのですが、日本は提供元基準というのを取っているので手術をした病院側で容易照合が可能なのであれば、個人情報の一部のデータを出す場合も同意が必要だという規律でそれを十分に遵守していなかったというのが問題でして、行為自体が個人情報なのかというよりも、行為を記録した情報とか行為を記録した動画といったデータが個人情報なのかという議論に法律上はなると思います。

○村垣委員 あと1点です。では画像などがついてなくて医者動作自体の情報というのは、個人情報には当たらないと考えてよいですね。

- 殿村委員 そうですね。逆にその医師の個人情報に当たるというケースかなと思います。
- 村垣委員 ありがとうございました。
- 鎮西委員 私が引っ掛かっているのは、「この法律において、「個人情報」とは、生存する個人に関する情報であって」という部分が既に引っ掛かってしまって、今のビデオの話もそうなのですが個人に関する情報というのはどのレベルまで考えられるのだろうか。例えば今のビデオの話もそれは個人に関する情報なのではないかというところをどういう基準で決めているのだろうかという辺りはいかがでしょうか。意味が通じているか分からないのですが照合容易性といった辺りに関しては大体分かってきたのですが、肝心の「生存する個人に関する情報であって」というところをどう解釈しているのだろうかを私自身は引っ掛かっております。
- 殿村委員 非常に広いというように考えていただいて、個人の顔といったものは当然ですし、例えば行動履歴といったものも個人に関する情報で例えば私が AMAZON で何を買っているかというのも私自身の肉体的な情報ではありませんが、私がこういうものを買ったということも個人に関する情報だと言われていきますし、私がこういうものが好きだということも個人に関する情報だと解釈されているので、かなり広いものだというようにお考えいただくとよいかなと思います。
- 鎮西委員 なるほど。中田先生が最初のほうに質問された合成された情報に関してどこまで引っ掛かってくるかという話があったのですが、例えば新聞記者が誰かの記事を書いた場合というのは、この記事は新聞の記事であってそれを個人情報だとは言わないですね。創作性があるのかどうかは分からないのですが、著作物に対してそこに出てくる人のことを個人情報だとは言わないような気がするのです。そのアナロジーでいった場合に、画像合成で出てきたものというのは顔だったら個人情報になってしまうのかというところはいかがでしょう。
- 殿村委員 合成したものというのが結果的に実在する人のものではないのであれば、保護対象になる個人はいないということになると思いますので、個人情報ではないということになるのかなと思います。ただ、AとBという人の画像を重ねて、その画像を見て明らかにAとBを重ね合わせていることが分かればAとBの個人情報になるということなのかなと思います。問題意識をうまく汲み取れていないのかもしれませんが。
- 伊藤副部長 先ほどの提供元基準について手術動画を例に挙げると、その動画が手術した病院の中でひも付けられないということは100%あり得ないと思うの

です。ですのでその病院の中では誰がいつ手術したというのは絶対に分かっていて、ただその動画を患者の同意の下で提供したときに例えばそこで匿名加工をしたと、私たちが手術動画のデータベースを作ったときに匿名加工を誰ができるのか、どこまでやれば匿名加工と言うのかというのは、次世代医療基盤法に基づくものとそうではないものがあるのではないかと理解していたのですが、そこはいかがでしょうか。

○殿村委員　まず本当に真摯な同意があればそもそも加工する必要が法律上はなくて、なぜ加工が法律上必要なのかと言うのもともと特定していた利用目的ではない利用目的に使う場合とか、一時的に取得した医療機関から第三者に提供することに対して同意が必要で、それらについて事前に真摯な同意を得ているのであれば法律上は加工をする必要はなくて、あとは安全のために念のため加工をして出すということになります。

他方、事前に真摯な同意を取っていない場合というのは法律上は同意がないと第三者に提供できないですし、同意がないともともとの利用目的を超えた利用はできないということになります。

例えば利用目的を超えた利用をしたいので加工をするということであれば、この加工は誰でもできますので病院自身もできますし提供先でもできます。ただ多くの場合は病院自身が匿名加工情報までを作るとするのが難しいので第三者に作ってもらいたいと。ただ第三者に作ってもらうにしても渡さなければいけないのでそういうところで同意が必要で、でも同意を取っていないと。そのときに要配慮個人情報でなければいわゆるオプトアウトというもので、事後的に嫌だという人だけを排除するオプトアウトという方式が使えるのですが、要配慮個人情報は使えない。ここをできるようにするというのが次世代医療基盤法というもので、一定の緩やかなオプトアウト制度に基づいて認定された加工業者にはオプトアウト的な制度で出せるようにする。ただ次世代医療基盤法に基づいた認定事業者がやる匿名加工というのもいわゆる普通の医療機関がやる匿名加工と基本的には同じで加工基準は大きく変わらないという理解なので、そこは同じ基準でやっているということだと思います。誰がやるのかで他の人にやってもらうのであれば、それは同意が必要なのかどうかということによってくるのかなと思います。

○伊藤副部長　分かりました。どうもありがとうございます。

○佐々木委員　教えていただきたいのが、病理学会という学会単位で多施設から学術研究目的で病理の顕微鏡の画像を収集いたしました。その際その画像というのは特殊な疾病ではなくて、例えば胃がんとか大腸がんのように顕微

鏡で見てしまえば画像だけを見ても個人が特定できないというようなものになっていて、ただし集めるときにはオプトアウトで集めています。いわゆる個人情報保護委員会が定めたような手順に従ってやったような匿名加工情報ではなくて、医学研究目的の旧倫理指針に沿った匿名化された情報の形で集めてデータベース構築を行いました。

恐らくいろいろな施設から集めてそれが特定されないような状況になっておりこのデータの活用の方法について教えていただきたいのですが、この情報というのは匿名加工情報に当たるかどうかということと、オプトインをしていないのでオプトアウトで例えばいろいろな施設で共同利用してこれを学術研究目的でまた更に使うことができるか、あるいは企業が入った商業目的でこれを使うことができるかどうかを教えてくださいたいのですが、よろしくお願いいたします。

○殿村委員

学術研究目的での例外というのは、ここの利用目的の例外にも入っていますし第三者提供の例外にも入っています。したがって学術研究という枠組みの中でデータを提供するということに関しては、法律の世界ではオプトアウトは要らないのです。オプトアウトと学術研究というのはある種並列しているそれぞれの例外なので、どちらかでよいというのは法律の世界の話です。

ただ従来の倫理指針というのは、学術研究の例外でやるときもオプトアウトのような手続でやりなさいということになっていて、それに従ってやられていることがあるのだと思います。それはある種の倫理指針上の要請も踏まえた対応をされているということだと思います。いずれにしても学術研究の例外という枠組みの中でやられているので、その枠の中である限りにおいては更に他の第三者に提供することも異なる学術研究目的で使うということもできます。

最後におっしゃった商業利用でということところは、学術研究という名前からしても学術研究ではないものは駄目ですので、商業利用には転用できないというのが原則論になります。

ただ時間もない中なのですが、未解決の問題の所に書いたのですが、学術研究の例外で同意なく出したその情報を仮名加工情報に加工することによって学術研究以外の目的に目的を変更して商業利用することを検討されている方がいます。各制度のいいところを組み合わせたみたいなやり方なのですが、これがよいのかというのは、よいと言っている方もいますが、私は疑義があるというか慎重にしたほうがよいのかなと思っています。その辺りは未解決の問題なのかなと思っています。

います。

○佐々木委員 分かりました。ということは例えばアカデミアと企業が一緒になってコンソーシアムを形成して、その中でもともと学術研究目的で集めたデータを共に活用しながら最終的には企業が入っていますので商業目的で使うようになるということは、先生のお考えではまだ難しいのではないかとということでしょうか。

○殿村委員 そうですね。そのままでは難しく、例えば医療機関で仮名加工情報にした上で共同利用という形にして企業に出せば企業は商業利用ができるという形で、別の枠組みで使えばできる余地はあるのかなとは思っていますが、学術研究で企業が受け取っているものをそのまま企業の手元で仮名加工情報にしてしまうことができるのかというのは疑義があるのかなと思っております。

○佐々木委員 例えば病理学会というアカデミアで集めて元の医療機関もたどれないような画像情報と、それに新書きも付いたような情報が手元にあるのですが、それを具体的に仮名加工情報にするというのはどういう手続をすればよろしいのでしょうか。

○殿村委員 仮名加工情報にすること自体は先ほど申し上げた3つの加工基準に従った情報は消していくということのできる、これが例えば画像だったり動画だったりしたときにどうする必要があるのかというのは一番冒頭に御質問いただいたとおりまだ不明確なところはあるのですが、法律上はこれに従った加工をすればできるというのが法律の整理です。

○佐々木委員 分かりました。ありがとうございました。

○田中委員 2点ございます。仮名加工情報の作り方に関してなのですが、先ほど先生が例に出されたID01が殿村さんというもので、このIDを丸ごと削除したら連結できない、戻れないというのはよく分かるのですが、ここには置換ということも書かれていて、置換の仕方ですが、我々はハッシュ化というのをやって無意味な数字に変えるという形でやっているのですが、要するに仮名加工情報であると認めていただける条件のようなものがあるのでしょうか。これが1点目です。

○殿村委員 仮名加工情報の置換はあまり難しくなくて、殿村をID01というように置き換えるというのも置換になります。ですのでハッシュ化をするというのも置換になります。

一方で匿名加工情報のほうは置換の定義自体は同じなのですが、匿名加工情報というのは定義上は元の個人情報のほうに戻れないようになっていくというのが定義に含意されているものになっているので、単なるハ

ッシュ化をただけですとたどれてしまう可能性があるので乱数を入れた上でハッシュ化をしなければいけないとかそういった細かな基準はあるのですが仮名加工情報についてはそういったものはございません。

○田中委員 ありがとうございます。匿名加工情報は何が問題かと言うと、先生もポイントの所に挙げておられますが、次世代医療基盤法の在り方というのがまだ非常に曖昧であるという点があります。加工業者がどういうことを誰がやるのかが我々にはなかなか伝わってこないということと、先ほど次世代医療基盤法でもオプトアウトでというお話がありましたが、あそこには「丁寧なオプトアウト」ということがわざわざ書かれていて通常我々が考えるようなオプトアウトでは駄目なのではないかという懸念もあるわけです。

 ここの検討のポイントの所で書かれている次世代医療基盤法の課題として、先生が課題として感じられているのは何かを教えていただければと思います。

○殿村委員 ここは議論がなされているところなのですが、1つは必ずしも法律の問題ではないのですが利活用が十分に進んでいないというところがあるかと思っています。それが構造上の問題なのか若しくは匿名加工までしてしまった情報というのが有用性に欠ける部分があるのかとか、その辺りは法律の問題だけではなくて端的に言うとな十分に利活用が進んでいないというのが課題だと思っています。ただその原因については幾つかのことが指摘されているということかなと思います。

○田中委員 この中で私たちが一番気になっているのはトップコーディングとボトムコーディングというのがあって、非常にレアなものは削除しろということが次世代医療基盤法において個人情報保護するという意味において書かれているのですが、こうなると非常にレアな疾患とかレアなものというのは、先ほどから先生がおっしゃっている、提供元からすると容易に個人が特定できる可能性があるということで引っ掛かってしまうという部分があるのです。その辺は今も議論はそう進んでいないという理解で間違いないでしょうか。

○殿村委員 そうですね。今まさに議論されているところで、次世代医療基盤法はまだ大きく進展があるという状況ではないですね。

○田中委員 今我々はどうしているかと言いますと、アルゴリズムを作るまでの段階というのは学術目的、つまり人に対する医学的研究倫理指針に基づいてオプトアウトでやって、製品ができる商用転用のめどが付いたあるいは可能性がある段階で、バリデーションにおいては完全に個別同意を得て

使用するというやり方をせざるを得ないということで、そういう形で我々は進めているのですが現状でもそういうやり方しかないという理解で間違っていないですか。

○殿村委員　　そうですね。法律にカチッと従ってやるのはおっしゃっていただいたような方法でただそこをもう少し何とかならないかというので、1つが今年から始まった仮名加工情報というものをうまく使えないかというところで、各医療機関、企業の方が試行錯誤しているのですが、この仮名加工情報については先ほど法律上は利用目的も自由に変更できるし共同利用で事実上第三者に提供できると申し上げたのですが、今まさに厚労省で医療情報に仮名加工情報を適用するというものについて検討会がなされて考え方が示されています。9月末付けで一定の整理が出されていて、何らかの本人の同意だったり第三者機関による一定の審査がないと勝手に仮名加工情報に加工すれば何でもできるというのは医療情報についてはよろしくないのではないかという議論もなされているので、そこの議論の結果によってこの仮名加工情報がどのぐらい使い勝手がよいものなのか、維持できるのかというのは決まってくるかなと思っております。

○田中委員　　ということはこの仮名加工情報というのも法律の適用の中で考えられる考え方なので、遡及はされないという理解でよいですね。

○殿村委員　　遡及はされないで、仮名加工できるのは今年の4月1日以降です。ただ4月1日より前に取得しているものでも、今年の4月1日以降に仮名加工情報に加工することはできます。

○田中委員　　そういう立て付けなのですね。よく分かりました。ありがとうございました。

○佐久間部会長　　チャットのほうに2つほど出ていますが、個人情報でない仮名加工情報と個人情報である仮名加工情報の違いを具体的な例で御説明いただけますかということがあります。これは回答できますでしょうか。

○殿村委員　　先ほども議論に出たところなので右手と左手で申し上げますと、私の右手にはID01は殿村であるという情報があります。私の左手にはID01は殿村であり男であるという情報があります。私の左手にある情報から殿村というのを消す。これが個人を識別できる記号等なのでこれを消すと、ID01は男ですという情報になります。ID01は男ですというのは特定の個人、誰のことを言っているのかは分からないので、これは個人情報ではありません。ただもともと個人情報だったものから特定の個人を識別できる記号を削除して加工して作ったものになるので、仮名加工情報になります。

ただ ID01 というのは残っているのですが、私の右手にある ID01 は殿村だという情報と照合することによって ID01 の男は殿村であるということは分かるので、個人情報である仮名加工情報というものになります。更に ID まで消していたのであれば照合できなくなっていますので、個人情報ではない仮名加工情報というようになります。

○佐久間部会長 幾つか出ていますが2つほど。中岡先生から、同意なく学術目的で取得した要配慮個人情報を仮名加工することはできないのではないのでしょうかということですが、これはいかがでしょうか。

○殿村委員 仮名加工情報の大元が要配慮個人情報の場合ですと患者の同意が必要だと思っておりますと、これは要配慮個人情報を仮名加工するには同意が必要ではないかという御趣旨かと思えます。ダブルチェックをしますけれども、要配慮個人情報を取得するときにはもちろん同意が必要なわけですが、その取得している要配慮個人情報を更に仮名加工情報にするときには同意は必要ないというように整理されていますので、取得のときに例外を使っている場合にどうかということですね。

それは更に捻りが入るのですが、私の理解だと例外を満たしている以上は同意があったというように扱われることになるので、同意がなくてもそのまま仮名加工することができるのではないかと思うのですが、少し確認させてください。

○佐久間部会長 ありがとうございます。最後です。森先生から、いわゆる個人情報の定義要件の1つに「容易」という言葉が出てくるのですがこれはどのレベルで容易なのか。技術進歩があると容易さは結構変わってくるということをおっしゃっているのではないかと思います。その辺りは法律ではどのように考えるのかという質問ではないでしょうか。

○殿村委員 ここも非常に悩ましいところでして技術的に容易かどうかという問題と、例えば同じ機関の中に A データベース、B データベースというものがあったときに例えば機関内の規定で A と B は絶対に照合してはいけないとか担当者を完全に分けているというときに、技術的には誰かがやればできるのだけれどもがちがちにできないようになっているときに、容易照合性を切れるのかという論点がありまして、一応ケースによっては容易照合性がないというように言える場合もありますし、他方で言うと A 病院、B 病院という別の病院でもそこが技術的には照合できて例えば禁止などがされていない場合には、法人格が別であっても容易照合性があると言われる場合はあるので、技術的な容易性と、あとは契約とか社内規程などでどれぐらい制約されているのかを総合考慮して決めるという形になり

ます。

- 佐久間部会長 皆様の活発な御議論をありがとうございました。なかなか議論が尽きないところだと思うのですが、殿村先生ありがとうございました。

<検討の方針について>

- 佐久間部会長 皆さん非常に悩まれているところなので丁寧な御説明ありがとうございました。第1回専門部会で説明しましたように、当専門部会の目的は審査上の3つの論点に対する一定の考え方を示すことです。第2回の専門部会では報告書の項目と執筆分担について検討しました。今回はWG参加委員に各項目の概要を執筆いただきましたので、御説明いただきたいと思います。可能な限り簡便に重要なポイントを御説明いただければと思います。

まず海外状況としてIECとかISOといったところを含めて、中岡副部会長から御説明いただきます。同資料は会員限定の閲覧のため共有できませんので、概要説明になります。

- 中岡副部会長 もともとSNAIGのレポートのほうからということでしたが、若干衛研医療機器部のほうにもいろいろタスクをいただいています、海外の実態を調査していることもありますので少しその辺を補足しました。

メインのところはざっと読んでいただければよいかと思うので、皆様御存知のことが多いと思うのですが、まず一番大事な所はFDAに関してはかなり活発にいろいろな活動をしています。皆さん御存知のGMLPであるとか、その辺りの提案が今出されているというところです。

もう1点がPre-certですね。Pre-certificationプログラムの試験的導入、企業レベルで認証といったものを与えて、そこが出す製品であれば承認を簡略化するというような目的だったかと思います。これに関してつい先日の2022年9月にそのパイロットプログラムの成果物が公表されました。実はここに結論として書いてあることというのは取りまとめて言いますと、Pre-Certをやったけれども審査するときそれを反映させるためにはまだ検討が必要ですよというようなことしか書いていません。事実、Executive summaryにおいて試験的運用は完了したけれどもまだいろいろな課題が出たので、更なる規制枠組みの開発を継続しますと書いてあります。ですのでこれに関してもまだはっきりした成果が出ていないと読めます。

あと一応FDAはCDRH直下にDigital Health Center of Excellenceというものを設立しているのですが、実はこの内容というか活動実態が

よく分かりません。それに関しては場合によっては FDA の関係者に聞いてみようかなと思っています。またいわゆる ML を使用した SaMD に特化したような製品コードなどがあるのかを調べていただいたのですが、実はない。いわゆる従来の SaMD と ML-based SaMD とを区別する手段がないということで、実態が分からない。主機能に ML を利用した SaMD がどの程度承認されたかというのは分からないというのが現状です。日本はまだその辺ははっきりしているかなと思います。

その後も実際 FDA の話にはなるのですが、例のいわゆる GMLP ですね、Good Machine Learning Practice に関しての方針を示す文書を実はカナダと英国とで共同で発行しています。それに関してはこれを基本として IMDRF とか ISO/IEC、各学会とは議論を展開してくださいという形になっています。細かいところは省きますけれども、実際にこれを本当にできるのかというところは、なかなか個人的には疑問を持っています。

IMDRF のほうに行きます。IMDRF の実態に関してはもう既に皆さんいろいろな所で話を聞いていると思うので、これは特に細かく言うことはないかと思いますが、2022年5月6日に用語のガイダンス文書が発行されました。今後もワーキンググループ内で議論を行っていきます。先ほど言った GMLP とか PCCP、事前変更計画ですね。FDA が基本的には提案しているものですが、これも 2025 年までに検討しようということになっています。ですのでこの辺りにしてもまだまだ社会実装に向けて活動しているだけで実装しているところはほとんどありません。

国際基準、国際標準のほうに関しては、先ほど言いましたこの文書を作った SNAIG、Software Network and Artificial Intelligence Advisory Group がいろいろ調査を続けています。彼らは今、規格に関しては基本的に 3 層構造として活用しようと呼びかけています。基礎規格、いわゆる医療機器として要求される各種規格にプラスして、AI の基礎となる要求事項、包括的要求事項を記載した規格というものを作ろうと。それと 2 番目に、機能に応じた機能規格、AI functional な規格を作りましょうと。さらに意図した用途に応じた手法やその評価に特化した規格、AI particular を上乗せするような形で、実際に 3 層構造の規格を作ってそれで評価しようというところなんです。

今チャートはできましたけれどもまだ調査はしているのですが、現時点で PCCP を使って承認されたような SaMD は私どものほうでは確認していません。今も調査実行中ですのでもう少しお待ちいただければよいかと思いますが、現時点では分かっておりません。

AI base、AI functional 国際規格というのが今存在しませんので、こういうものを作りましょうということをSNAIGが提案しています。これはどこでどう使う、どこのTechnical Committeeで作るのかといことに関しても多分これから議論されていくのだと思いますが、この辺りは少しウオッチしていく必要があると思います。雑駁ですけれども現時点での分かっていることをまとめた資料の説明は以上になります。

- 佐久間部会長 ありがとうございます。今の点について何か御意見がありますでしょうか。私もここで書かれているGood Machine Learning Practiceの10項目、原文を読むと結構これはどういう意味だということがいろいろ書かれています。中岡先生、その辺りはもう公表されている文書なので、GMLPについて少し解説をしてその背景などをレポートに入れていったほうがよいでしょうか。できるかできないかという議論は確かにあるのですけれども。
- 中岡副部会長 そうですね、彼らが言っていることは何となく分かるのですけれども、具体的にこれが何なのか説明せよと言われると、なかなか難しい。
- 佐久間部会長 そうですね。
- 中岡副部会長 他のものを読んでいてもそうなのですが、割と概念的なことを言うことが多くて、特によく言う透明性が大事といったものなのですけれども、では透明性とここで言うのは何の透明性なのだというのが少し分かりにくいですね。
- 佐久間部会長 読んでいて、実はこの議論でやっているバイアスの問題だとか再学習の上での課題だとか、結構この10項目はマッピングできるような気がしています。その辺りは関連付けて説明していくと、これのある意味はこういうことではないかということが書いていくかなと思ってしまして、その辺りは議論をさせていただければと思います。
- 中岡副部会長 ありがとうございます。
- 村垣委員 1点、こちらで先生も今話題にされているAI baseとAI functional国際規格に関して、国内の動きというか、国内委員だとかコミナーとかその辺のところの国関係の今の動きはどうなっているのでしょうか。この3つを立てている主要国はどこで、日本も委員として入っているのか、その辺の情報はありますか。
- 中岡副部会長 そこは佐久間先生のほうが詳しいかと思います。
- 佐久間部会長 SNAIGには日本から入っています。これから立ち上がるのですけれども実はAI本体、これは医療用ではなくて一般的な産業用のものについても既にいろいろな議論がされているようで、そこには具体的にどなたが入

っているか分かりませんが、これから立ち上げるに際して手を挙げている形になるのだと思います。ただ本来ならそこを進めているような方が行っていますので、ヨーロッパ、アメリカ。ヨーロッパ主導という形になっていくのかなと感じています。

○中岡副部長 そうですね。

○佐久間部長 ISOもそうですね。

○中岡副部長 基本、SNAIGが今活発に提案を取りまとめているという印象で、個別のTC、ISO、IECでのTCが積極的に動いているという印象はまだないです。ただ唯一評価に関しては確かIECのほうで提案が出ていまして、それが成立するか否かでこれから先動いていく可能性はありますね。

○村垣委員 そのこのところは本委員会とも相当関わってくるかと思ったので質問しました。

○中岡副部長 はい、その通りだと思います。いろいろ情報が入り次第、皆さんと共有したいと思います。

○村垣委員 ありがとうございます。

○佐久間部長 他にございますでしょうか。では次にバイアスについて第2回で御説明いただいておりますが、森先生の資料を事務局のほうで出せますか。

○森委員 やはりバイアスについては幾つかの問題点がありまして、それに関してはまず書いているとおりです。その中でもデータについては、例えば病院によって出てくるバイアスであるとかあるいは本当に手法的なデータのセレクションとかで生じるようなバイアスにも幾つかのものがありません。それを整理して書きました。

それから後は利用手法が持っているバイアスです。これは手法的なバイアスという形で、その手法を使うとどうしても避けられないバイアスが生じたりしてきますので、そういったバイアスを現在取りまとめているところです。

それから人によるバイアス、人が使っている間にラーニングカーブなど、生じてくるようなバイアス、使う人によって生じてくるバイアス、あるいは設計する人によって生じるようなバイアスというのも絶対にありますが、そういったこともコメントとして付け加えていく予定です。

○佐久間部長 ありがとうございます。先ほど Good Machine Learning Practice、GMLPの所で透明性だとかを言っているのが、この辺りと関連しているように思います。使い方のバイアスというところですね。これがどういうものかきちんとしないと、使うほうも変な使い方をしてバイアスしてしまうといったことかなと思うのですけれども、透明性に関して言うところ

の辺りは何か関係ありそうですか。

○森委員 透明性は幾つかの問題点が、使う方もそうですしデータのソースの透明性というのもあるでしょうし、一般的な話なのですけれども、非常に偏ったデータを出してきているデータソースとかもあったりするので、そこに例えば透明性があるかどうかということがまた問題となってくるわけです。透明性という観点も、今回挙げました幾つかの事項の中の全てに関わってくるのではないかと思います。これは実をいうと結構いろいろなことが全て関わってきているので、組み合わせを考えるともう爆発するような状態かと思えます。

○佐久間部会長 ありがとうございます。他に何かありますでしょうか。こういう形でバイアスと一言でいろいろな文書に出てくるのですけれども、そのバイアスにどういうものがあるか少し整理して、論点整理していくと、多分ここの専門部会としての1つのアウトプットになるかなと思います。ありがとうございます。

○鎮西委員 透明性の話になるのですが、基本的に企業はその辺り、透明性ということと企業活動というのはあまり相容れないところがあると思います。そうなってくると逆に透明性を担保するのは、そういうデータを見る権限がある PMDA ということになってくると思うので、透明性の話を書くときはそういう観点というか、要するに企業は本質的に企業努力で集めたデータについて開示する義務を負っていないし、でもそうは言いながらどこまでを社会に対して役すべきか、あるいは PMDA に代理になっていただくのかという辺りの話も必要になってくるかと思えます。そこまで踏み込んで書けるかどうか分からないですけれども。

○佐久間部会長 今の透明性の話は恐らくデータ本体を開示するのかということとは少し違って、多分データの属性などについて明らかにしようということが、まず第一段階として出てくるのだらうなと思います。その辺りをどう明らかにしていくのか、どこから取ったのかとか。意外になかなか出しにくいところもあるかもしれませんけれども、その辺り、今、鎮西先生がおっしゃったところは少し書き方を検討すべき点かなと思います。

○鎮西委員 あるいは企業にとって出しにくいところを PMDA はきちんと確認しておかないといけないという点もあるのかもしれない。

○佐久間部会長 そうですね。ありがとうございます。それでは次に行きたいと思えます。次は再利用と市販後学習のポイントについてです。第2回で御講演いただいた清水先生をお願いしてまとめていただきました。清水先生、お願いいたします。

○清水委員

この報告書の詳細は後で御覧いただければと思いますが、幾つかの部分からなります。まず最初に再学習の性能評価の必要性。それから想定される評価法の例を短く書いています。その後で本題である同一のテストデータを繰り返し用いて評価を行う場合にリスクのある再学習法の例とその対処法、それからコンペティションのデータを用いてリスク評価を行った例についてまとめてあります。

2節が性能評価の必要性、3節が評価法、性能の評価法の例を書いてありますが、詳細は御覧ください。4節がDworkの論文、それからGossmannの論文をベースに書いてあります。今日は詳細は説明しませんがこの図を簡単に使って説明しますと、本来は50%の分類精度が出る段階であっても同一テストデータを繰り返し利用する際に分類器の設計にテストデータの結果を使ってやると、50%からどんどん離れていってしまうというバイアスが生じてしまうという話です。

これはGossmannが提案しているそれを補正するためのアルゴリズムで、コアとなる部分がこの2の(A)の所で、この条件を満たすときのみテストデータにノイズを加えてアクセス可能という方法で、それ以外はトレーニングのデータを渡すという方法となっています。Gossmannらのアルゴリズムを下に書いてありますが、医用画像でよく使われているAUCをベースにした方法、あとは基本的には同じものですので割愛します。

5節のほうで、Kaggleでの同一テストデータによる評価とそのリスクの評価。これも前回説明したものと同じになりますのでそこは図を御覧いただいて簡単な説明をして終わろうと思いますが、これは横軸がプログラムの開発者がコンペティション当日まで繰り返し見ることができるテストデータに対する評価値です。このPublic accuracyを見ながら繰り返しシステムをアップデートして、コンペティション当日に備える。コンペティション当日にはプライベートなテストデータの結果が公表される。この両者に差がなければ、同一テストデータを繰り返し利用してもそれほどリスクがあるわけではないということになります。

これは代表的な4つのコンペティションについての結果と、その4つのコンペティションのトップ10%の結果です。中にはこのように少しpublicのaccuracy、つまり当日まで繰り返し見ることによる性能のほうに偏ってしまう、テストデータを繰り返すことによってそのバイアスが入ってしまう傾向が見られましたが、数値としてはそんなに大きな差ではないということで彼らは120のコンペティションのデータを解析して結論としてはそんな大きな問題は見られなかったと言っています。

ただ例外が幾つかあって、その例外が起こる条件がここです。データセットの分割が不適切で2つのデータセットが独立同一分布ではないという場合とか、そもそもテストデータの数が少ないという場合には、少し偏った結果が起こりうるというようなことも注意として指摘しています。

これは前回までに報告しなかった新しい情報として今回付け加えたのですが、彼らはデータセットのサイズとそれからこの2つの public と private の差の関係も調べていて、大体 1,000 とか 1 万ぐらいになると両者の差が小さくなる傾向とも報告しています。こういうことを基に今後同一テストデータを繰り返し利用する際の問題点とか、どのようにすればバイアスが含まれるリスクを小さくできるかという議論を進めていけばよいのではないかと考えています。

最後のほう、ここを後で御覧いただければと思いますが、以上を踏まえてどのようにすればよいのか少し案を考えてみましたが、これが全て完全にバイアスをなくすような方法ではないので、こういうことを御覧いただいて今後の議論の足しにさせていただければと思います。私からは以上となります。

○佐久間部会長 ありがとうございます。御質問ございますか。前回お話いただいたことを更に補填いただいて最後に評価上のこういう改善の方向性もあるやもしれないという、これは非常に学術的にはフロンティアというかまだこういう議論が進んでいるところなのでなかなか明確なことは言えませんが、方向性としてはこういう議論もあるということかなと思いました。

清水先生からここまで詳細なことを書いていただいて、さらにこれは最後の段階でどうなるのでしょうかという御説明がありましたけれども、更に環境とかそういうことがあって、また委員の中で議論等があればそれを議論させていただければと思います。ありがとうございました。

あとこれに対して今回の原稿では論文からの図をふんだんに取り入れていただいたということですが、実はこれは著作権法上の問題があって版權に抵触するおそれのある図表は報告書に出さないというところの問題もあって、最終的におまとめのところでライセンス保持者から許諾が取れない限り出せないということで、通常のレビュー論文を書くときと多分似たような形の手続ですが、そういう点を御配慮いただければと思います。

○清水委員 分かりました。

○佐久間部会長 次に「データ augmentation の新たな手法」について、中田典生先生、御説明をお願いいたします。

○中田(典生)委員 データ augmentation について簡略にしか作成できませんでしたけれども、最初の節、パラグラフで簡単な GAN の説明をしてあります。これは今回、時間の関係で省略します。

2 番目のパラグラフで、現在の Pubmed での検索状況の検索結果を簡単にまとめてあります。これも日々増えているのでまだ足りない部分もあると思いますけれども、取りあえず GAN、医用画像を使ったものについての論文を検索してまとめました。そこに書いてあるとおりの胸部単純写真、胸部あるいは腹部の CT というのが結構多くて、次に頭部の MRI、病理組織像、これら複数の論文があります。それから単独、1 つずつですが、皮膚病変の湿疹、目で見えるインスペクションの画像、マンモグラフィ、歯科領域のセファログラム、歯の写真、MRI-PET 画像、こういうようなものに GAN が使われています。

GAN を使うに当たって、まだ今は GAN、合成画像をトレーニング用の画像として使うものしかないということ。それから半分以上は、GAN は画像の合成、要するに画質の向上に役立てるとというのが結構出てきていますが、最近はトレーニングに使うほうがだんだん増えている状況です。最後にサイノグラムとか、raw data のケーススペースの話を書きましたけれども、これは現在進んでいるという状況なのでそのことについて触れておいただけなので、場合によっては正式な文書を書く場合、これを除いてももう少し GAN の細かい論文の検索の結果について書いてもよいのかと思います。以上が、私からのまとめの報告です。

○佐久間部会長 ありがとうございます。中田先生が今おっしゃったこの画像生成というのを、いわゆるモデルの設計の段階、トレーニングの段階で使うということが増えてきていると。一方で評価に関しては、そこは実際のものを使っているということなので、設計方法の検証みたいところでこういうものが出てくる可能性があるということかなと思いますが、そのような理解でしょうか。

○中田(典生)委員 そうですね。今回は私の私見はなるべく一番最後の所以外は書かなかったのですが、先ほど個人情報の問題でも話したとおりの、全て合成画像によって置き換えた画像が本物の画像とクオリティが同じなのかどうかを比較検討してみたことがあるのですが、今の段階では同等の特徴量を得ることはできなさそうなのですが、新しい GAN の手法がどんどん出ていまして今後それを検証するかどうか。もしそれでほぼ同等となると話が違ってきますので、個人情報を置き換えるような個人情報ではない画像群みたいなのを作れる可能性も無きにしも非ずなので、引き

続き注視すべきかなと思います。

○佐久間部会長 ありがとうございます。多分先ほどの Good Machine Learning Practice のところで評価データのことは今分かりましたけれども、訓練データについての独立性だとか、そういうことを求めているところがあったときに、こういうものが使われてくる可能性があるのだろうと思っていて、その辺りを今御指摘いただいた、現状ではまだまだ同一性というところで限界があるかもしれないけれども今後その同一性が高いものが出てくるといったときに、非常に医療情報が数が限られているときにこういう augmentation の方法を使うことが出てくる可能性があるということかと思っておりますので、そういう点で書いていただくということかと思いました。

ではまた近いことになるかもしれませんが「物理モデル・シミュレーションによる学習データ構築の現状と課題」です。現実にはないものを物理モデルを作りながらやっていくということで、データ augmentation の1つの手法となるかと思っておりますが、それでは鎮西先生お願いいたします。

○鎮西委員 数値シミュレーションの報告書を2年ぐらい前に書いてありますので、それを引用する形で書きましたが、数値計算だけではなくて数値計算と機械学習を組み合わせて学習させました、あるいはそのものを医療機器にしていくといったものも将来的には予想されるでしょうということをお願いいたします。

機械学習で数値シミュレーションを使うとしたらどのようなものがあるかということで、これは多分に想像というかこれだけでよいのでしょうかというところに関しては十分に調べられているか分かりません。ただ私が調べた限りでは、まず1)のデータ augmentation のためにデータを数値計算を使って生み出してというのは、その数値計算がどのようなものであるかによってくるわけですが、かなり昔からあったような気がします。逆に、では例として何を引くかというところはお知恵を頂ければと思います。

2)番目のほうは、これは例えばマシーン・ラーニングと finite element とかで検索して出てきたものを調べてみて、ある程度まとまったものになっているのが引用として書いてある2つくらいでした。そのうちの1つが顔面再建とかをした後で表情筋が例えば左右で均等に動かなければおかしい、あるいは自然なスマイルをつくるのにどういう訓練をするかというのがあって、その訓練などの目的でいずれ使いたいという、そういうネットワークを育てる話なのですが、全部をネットワークでやるので

はなくて筋肉それぞれを動かして on・off とかで規定してやって、それに対してでは最終的にどのような表情になるかという、表情を生成する部分を計算でやっていますと、そういう例です。ですからハイブリッドという感じです。その計算した結果を、またこれを使って discriminator とかを調教するというをやっている例でした。出てきたのはこれくらいで、あと似たような形のリハビリテーション用というのも、もう1つか2つくらいできたような気がします。広い意味でメディカルなのでここで引用してあります。メリットとかのところを書いてありますけれども、この辺りから先はほぼ私の意見になってしまっているのをこれを載せるかどうかについて御意見を頂ければと思います。

まずメリットとして想定されるのが、今我々が持っている数値シミュレーションのモデルは物理モデルばかりなので、その他の生理学的なモデルあるいはメンタルのモデルというのはほぼないです。機械学習に逆にそういう部分を任せてしまうことが可能になるという、最初の例は先ほどの顔面再建の話は正にその例だと思います。データ augmentation にしろ今の例にしろ、数値シミュレーション部分というのはほぼ順問題で記述されているので順問題、静定的な問題という感じなので、数値計算としてはやりやすい。複数の状態方程式とか出てきてもそれを解くこともできるだろうということで、数値計算だけでやろうと思うと大変な問題がかなり扱いやすい問題になるので、逆によく分からないところをネットワークに押し付けているということかもしれません。

あとは医療の場合、必要なパラメータが患者から得られない又は代表値からしか得られないという状況が多いという話題を、前の数値シミュレーションの報告書でも扱いました。これもまたそういう曖昧さの影響みたいなのをネットワーク側に押し付けているやり方かもしれないです。これがよいのかどうか分かりませんが、そういう使い方ができているということです。

逆に留意事項としていろいろ書いてあります。まず2つ組み合わせるのうまくいかなくなったときに、どちらが悪いのだということ、リバーエンジニアリングというか、それは難しくなりますよという。それから数値シミュレーションに関する FDA 公認の ASTM 規格で V&V40 というのがあるのですが、その中でもネットワークとか組み合わせってきたときには、uncertainty quantification、その不確かさを表すパラメータの導出とかの手法はそもそもありませんと言っています。

後はバイアスの問題に関して言うと今度は逆に数値シミュレーションだ

けでデータを作っていく、数値シミュレーションで表現できるモデルばかり調教する形になるので、そこは問題があるかもしれません。特に数値シミュレーションには不得意分野があって、病変であるとか、そういった例外的な事項、あるいは特異・非線形であったり、過渡応答や非定常であったりとか、そういうのは単純化せざるを得ないのでうまくいきませんという話です。

それから下から2番目のポツで、必ずしも演算量が減るとは限りませんということ、メモリーの浪費になるかもしれないということ。特にメッシュのサイズを変えていくことになると、別の問題を生む可能性がありますということです。

後はデータ augmentation で使うデータというのは、アフィン変換とかをやっているものに関してはその医学的な意味というのは特に問題にすることはないわけですが、では同じようにその考え方で当てて数値計算で生んだデータに関して医学的妥当性とかに関して別にバリデーションは必要ないですよという言い方ができるかどうかはケース・バイ・ケースとしか言いようがないというまとめ方です。

最後に、これらに鑑みるとまだまだなので数値計算を使うのであったら、健常の状態の学習に使うであるとか、数値シミュレーションと臨床でデータを取ってきてデータを学習させる、あるいはネットワークを鍛えるといった場合の双方の得意分野を組み合わせる使い方がよいのではないのでしょうかというまとめにしました。

- 佐久間部会長 ありがとうございます。それではこれまでのところで御質問や御意見はございませんでしょうか。
- 鎮西委員 私の原稿の中で黄色く書いてある辺りは、基本的に見ていただければという意味です。
- 佐久間部会長 画像分野では結構古くから学習に使われているのではないかということがあって、その辺り何か適切な例があれば後ほど鎮西先生あるいは事務局に投げただけければ、取り入れることができると思います。
- 森委員 鎮西先生ありがとうございます。思い出しました。昔、例えば肺がんの埋め込みでスペキュラありなどといって、何かギザギザが付いた僧帽形の立派なものの三次元バージョンを確か今、山口の平野先生が作っていて画像に埋め込んで実験をしたとか、確かに例が幾つかありました。ありがとうございます。
- 佐久間部会長 ありがとうございます。それは研究レベルではされているところがあって、一方で多分そのモデルを鍛える段階ではあるのかもしれませんが

ども、評価の段階ではどうなるのかというところは次の課題になるのかと思います。

一方でこれは前回の数値シミュレーションのところでAIが使われてくる、ちょっと逆方向の議論ですけれども、そういうアプリケーションがあることを少し御指摘いただいたのかと思っております。他に鎮西先生、委員の方に聞きたいこととかあれば。今の点、森先生から御回答がありましたけれども。

○鎮西委員 あるとすると、この1)と2)で、1)はデータ・オーギュメンテーションなので、データをたくさん作ってやって、学習させてという話だと思います。あと全然扱わなかった話題として、ではそういうデータをテスト用のデータとして使えるか、あるいはバリデーションデータ、validationというのは、深層学習の上でのvalidationで使えるかどうかというのは出てこなかったもので、何も書いていないです。

○佐久間部会長 今のところはやはりまだそのトレーニングの段階というのが多分現状かなということですが、一方で動物実験が制限されるといったことが出てきたときに、何か出てくる可能性はないわけではないということかもしれません。

○鎮西委員 はい。

○佐久間部会長 引き続きまたよろしく願いいたします。それでは「データベースについての展望」について、お話しいただければと思います。伊藤先生、笹野先生、田中先生、佐々木先生の順でお願いできればと思います。

○伊藤副部会長 我々は手術動画のデータベースの構築ということで、AMED様の御支援で3年間に掛けて手術動画を構築してまいりました。その結果、内視鏡外科学会と学会のサポートもあったのですけれども、13術式、広い術式で4,000を超える手術動画を収集してデータベースを作りました。

研究機関から頂いた患者情報、術者情報とともにこのデータベースに格納していて、今現在AMEDが終了しましたがけれども現在もデータベースを維持しています。収集活動も継続していて、2021年11月最近ではデータベースの利活用はまた企業にもそれが及んでいるということを書かせていただいております。

手術の動画のデータベースにおける生じた課題と解決というところで、持続可能な収集体制、これは非常に大きな問題だと思います。我々は基本的にはAMED様のが終わりました、データベースを持続運営する Surg storage という会社を立てて、彼らによって持続運営をお願いしているという状況に行いました。

動画の扱いですが、これは先ほどの議論と少し外れているのかもしれませんが、我々、特徴は、学術利用とともに、その企業の産業利用も全て最初のオプトインでお願いする形でデータベースとして使いますよと言って患者さんをお願いしております。ですので丁寧な形でお願いしておりますが、我々の中で仮名加工情報にしまして、その後 Surg storage 社の中では、更に対応表を削除したり付帯情報をマスキングすることによって、当方の弁護士や個人情報保護委員会などとも相談しながら匿名加工化して産業利用に使っていますよということを書いております。問題だったのが、その手術動画の権利の帰属だとか、そういった中では同意の中で各人の権利を放棄していただくことを明記して収集しております。患者さんの同意に関しては先ほども書いております。

動画提供における作業においてはこれはクラウドでも作業をしたかったのですが、各施設の IRB の時間が掛かったりするというところでなかなかしにくいところもありました。収集する動画の規格の統一化や付帯情報の確定については、まだ拡張子の変更や動画出力時の圧縮の画質の変更等、まだまだやらなければいけないことがあると感じています。動画の管理場所については最初はクラウドを想定しておりましたが、今現在では費用のことも考えてローカルサイバーの中で行っていることも書いております。

将来の課題及び将来展望においては3つ書いてありますが、収集スキームの効率化、これはなかなか大きな問題で、これは時間もありますので、書いておくだけにします。手術動画の規格の標準化及びユーザー側、データを大規模に利活用する企業、病院をいかにエンハンスしていった、その産業の導出、いわゆるデバイスまでつなげていくかということかと思っております。以上です。

○佐久間部会長 ありがとうございます。これについて御質問はございますでしょうか。現状をまとめていただいて、今日の個人情報の議論などもあって、その辺りの議論がまた少しアップデートされるのかなと感じられました。

○伊藤副部会長 そうですね。ありがとうございます。

○佐久間部会長 それでは笹野先生、次にお願ひできますでしょうか。

○笹野委員 私は前回の資料からのアップデートは今日お見せするものはございません。今日の議論を踏まえて心電図データに関してまとめさせていただきたいと思います。他のデータと少し違うところというのが、教師データが100%ではないと。陽性にしても陰性にしても100%ではないということなどを中心にとということで、特に心電図より予測のところに向けた、

疾患発症予測等に向けたデータベース構築についてまとめさせていただきたいと思います。

○佐久間部会長 ありがとうございます。それでは田中先生、お願いできますでしょうか。

○田中委員 あくまで私たちが大きく異なるのは、先ほどありました要配慮個人情報たるテキストデータと画像をバインディングしているところを書いているところです。もう1つは標準化フォーマットとしてアノテーションをする際に各研究者がばらばらのことをするとデータベースとしてうまく保管ができないので、消化器内視鏡においては統一フォーマットで集めるということを記載しました。

またいろいろ御意見を頂きたいと思いますが、書き方のフォーマットがどうも私よく分からなくて、先生方からいろいろな御批判とかこのように書けという御指導をいただければ、また改めさせていただきたいと思っております。

○佐久間部会長 ありがとうございます。それでは佐々木先生、お願いできますでしょうか。

○佐々木委員 私の場合もやはりAMEDの支援で病理学会として画像を集めたのですが、今日の殿村先生のお話であったように集めたデータをどのように公開していくのか、活用していくのかがよく分からなかったのです。結局いろいろAIのプログラムの活用とかを書かせていただいたのですが、今日分かったのが旧倫理指針で匿名加工情報、匿名化された情報として集めたそのデータも4月1日以降に仮名加工情報に変えると活用できるというお話があったので、1つはデータの利活用については今現在アーカイブ化した病理画像が20万枚ぐらいありまして、それに対して9万枚ぐらいはアーカイブ、会員に限定して公開をし、学術研究目的で使えるようにして、今後研究者の方に提供しようという流れで書かせていただいたのですが、今日のお話でそうはいってもコンソーシアムを組んでもまだ商業ベースは難しいということをお聞きしましたので、その辺のところも書き込んでまとめたいと思います。

ただやはり書きぶりというかフォーマット、田中先生もおっしゃったようにどのように書いてよいのか後で御指示いただければ、それに合わせてまた直したいと思いますので、よろしくお願いたします。

○佐久間部会長 今、書きぶりをどうするかという議論だったと思うのですが、実はこの部分は少し悩んでいるところがあって、やはりデータベースを作るときに前半でGood Machine Learning Practiceという言葉が出てきていたと思いますが、少しその原文に戻って議論しなければいけないと思います。

あとバイアスがないかとか、その対象としている患者群の特性をしっかりと表現しているかどうかとか、実は結構実際の評価とか設計の検証というところでは議論になるところがあると思います。そういうところに対してこういうものを活用できるようにするにはどうするかは多分一つ大きな課題なのだと思います。その辺りを少し、今世界的にこういうデータに対して求められる品質ということに対して幾つか提言が、まだまだ抽象的な議論が多いのですが、それをこれまで組まれてきた先生方の中から見ていただいて振り返っていただいて、どうであったかとか、今後のデータベース構築に対するそういう観点からの提言のようなことを書いていただくとよいのかなということで、全体を見させていただいて思いました。その辺り、他の先生方はどうでしょうか。

○伊藤副部長 今の御指摘で言うと、今4つのデータベースにやはり共通している課題とかもあると思います。それぞれ出していただいた後にシェアできる部分、共通の課題としてそのデータベースの構築に対するジェネラルの課題は1つまとめて、もう1つはモダリティによって違う難しさだとか特徴というのもあるので、やはり出てきた中でばらばらのテーマ4つを1回見た中で、シェアできるものとばらばらのものをまとめ直したほうがよいのかと思いました。

○佐久間部長 そうですね。その辺りをワーキンググループ等で少し議論できるとよいかなと思いますが、いかがでしょうか。

少し曖昧な形で先生方に発注を出したところも実はあったのですけれども、こうやって出てくるとやはり全体が見えてきたかなと思っています。前半で世界的な動向として、やはりこういう製品を開発するときのデータの在り方、それからこれを評価するときのデータの在り方に対して、Good Machine Learning Practiceといった議論が出てきているのですが、まだそれがあまり明確な形で示されているわけではないので、そこを少しひもといっていく必要があるのかと思っています。今日お示しいただいたバイアスの分類であるとか、あるいは再学習、学習データを繰り返し使うことの中での懸念点、それがそれほど大きく出ること出ないことがあるといったこと。それから開発の段階で新たな技術としてデータ augmentation のいろいろな議論が出てきているという点。後半、こういう製品を開発する上で重要なデータベース作りのところで、やはり落としてはならない部分は何とか、そういうことについて議論することと、最後、今日殿村先生にお話いただいた法律ですがそれに準拠するというので、実は今日お話を伺って研究者側が少し抑制的に動いていた

ところも実はあるかなと思うところもあります。佐々木先生がおっしゃっていましたが、現状の過去のデータをどうやって活かすことができるのかといった方向性も出てきたのかなと思いますので、その辺りについてうまくまとめていけたらと思います。先生方から、何かその辺りの御提言、御意見はございますでしょうか。

今最後、伊藤先生からデータベースについての共通の課題を少しまとめながら、一方で、個別、モダリティ特有のことというのも、非常に典型的な応用例も出てきていると思いますので、うまく示せるとよいかなと思います。

それでは、御議論ありがとうございました。本日の議事は以上です。事務局から他に何かありますでしょうか。

<その他>

○事務局（澁岡先端技術評価業務調整役） 次回の専門部会は、2月20日(月)の14～16時の開催を予定しております。詳細等につきましては、追って御連絡をいたします。また1月16日(月)の14～16時には、第3回のWGを予定しております。御参加いただく委員は、伊藤副部長、佐久間部長、佐々木委員、笹野委員、清水委員、田中委員、鎮西委員、中岡副部長、中田典生委員、森委員です。連絡事項は以上です。

○清水委員 佐久間先生、次回のワーキングの宿題は何か決まっているところがありますか。

○佐久間部長 今やっているまとめのところを少しブラッシュアップしていただければよいかと思います。あとワーキンググループで宿題として、先ほどのデータベースの共通課題とか、少し全体の構想を中岡先生にまとめていただいている、あそこのGood Machine Learning Practiceに対して先生方がやっている議論がどのようにマッピングできるかといったところを少し議論させていただければなと思っています。

<閉会>

○佐久間部長 では、本日の専門部会はここまでとさせていただきます。皆様、本日はありがとうございました。