

(2024 年 9 月 5 日プレゼン)

8.「医学領域における AI 活用の未来」

東海大学医学部講師／ブリガムアンドウィメンズホスピタル循環器内科講師 後藤信一氏

東海大学の総合内科と米国 Brigham and Women's Hospital の循環器内科で講師を兼務しています後藤です。本日「医学領域における AI 活用の未来」と題していますが、私の研究をメインに説明させていただきますので宜しくお願いします。

私自身は循環器内科のトレーニングをした医師ではありますが、何故 AI などをやっているのかと言うと、中学時代からソフトウェアの開発のような事を趣味でやっていたり、動画の管理ソフトや顕微鏡の画像処理ソフトを作成し売っていました。それが雑誌に紹介されるといった事があり、慶應義塾高校にこのプログラミングの一芸推薦で入り、入試を受けずに現在まで至っています。大学時代も高校時代も電子工学研究会、高校時代は電子工学研究会があったので、そこに入って部長を務めていましたが、大学では無かったので自分で設立し、今も存続しているサークルになっています。ロボットなどを作成して、相撲ロボット大会に出たりなどと、元々工学系にかなり興味があり、その中で父が医者であったという事もあり、医学の道も興味があり医学部に進み、できれば医学と工学を融合させたいという事で今の研究に至ったという経緯があります。

ここから本題ですが、最初に医療で活用されるデータを説明します。医療で活用されるデータというものは、それほど特別なものではないのですが、心電図や心エコーのような電位データの時系列、心エコーであれば動画ですので、画像に更に時間軸の次元が入っているような 3 次元データ。通常の血液検査のデータでも 1 回だけを見ると 1 つの数値ですが、通常医者が読む時には、前回はどのくらいで今、どのくらい上下しているかと時系列で見るので、この辺りを全てまとめて多次元データと捉える事ができます。

医師が多次元データを読めるようになるまでは結構訓練が必要です。訓練が大変だという事と、医師が全力で訓練しても必ずしも全部のデータが読めるわけではないという現状がありますので、これを専門医としての経験に基づいて読んでいくという作業が医師の専門性によってばらついてしまったり、状況によっては夜間の救急のような循環器の専門医がいないような状況ですと、心電図を撮ったが心筋梗塞を見逃してしまったというような事が起こるので、これを何とかしなければいけないというニーズが医療にはあります。患者さんが夜間の救急外来に来ると、見るのは当直医で研修医であったり、割と若手、または内科であっても通常内科当直であって、循環器内科というようにはなっていないので、非専門医が見るという事がままあります。この部分は AI がかなり活用できる状況かと考えています。

ここからは釈迦に説法かもしれませんが、AI がどのようなものかという事を簡単に説明したいと思い

ます。AI は非常に広いタームなので、AI と言うと何でも含んでしまいますが、我々が今、医療 AI と言う時には通常ニューラルネットワークによって構築される数学モデルというものを考えますので、今回の講演の中では AI とはこのニューラルネットワークを指すものとします。ニューラルネットワークとは、人間の神経細胞、動物などの神経細胞の機能を模した数学モデルという事になります。

ニューロンとはどのようなものか、非常に単純化して考えると、複数の樹状突起からいろいろな入力、刺激を受け取り、それを最終的に全部ひとまとめにし、それがある閾値を超えなければ何もせず、越えれば発火する。そして発火すると次に電気信号を、軸索を通じて出力するというような、かなりデジタルな働きをしているものと考えられます。これを数学的に表すという事で、どのようなになっているかと言うと、入力の数値に対してある重みをかけてやるという事を複数の入力に対してそれを全部足し合わせて、そして最後にこの閾値を超えた時に急激に発火するような機能を模した関数という事で、活性化関数というものを通す事によってニューロンの機能を表現しています。

このニューロンを大量に積み重ねて、膨大な量の計算を行うという事が AI でして、この AI とは比較的単純な計算を多数組み合わせるという事で成り立っています。この AI というものが、十分な数があり、ニューロンが複数層に重なっていると、どのような複雑な関数であっても近似できるという事が数学的に証明されています。ですので、ここに示しているようなハート型を描くような、これは結構複雑な関数だと思いますが、もっとずっと複雑な関数であっても先ほど申し上げた足し算、掛け算、活性化関数という、簡単な計算の組み合わせで、実はこれを表現する事ができます。コンピューターは単純な組み合わせの計算を何度も何度も間違えずに繰り返す事は得意なので、コンピューターにとって都合の良いシステムという事になります。

このような特徴を使うと、病気を分類するというような事ができるようになりますが、ここに簡単に例に示しているものは、赤い点と青い点があり、この点を 2 つにきちんと分けて下さいという問題があったとすると、線形分離と言われるような 1 本の直線を引くというような分け方ですと、どのような直線を引いても分ける事ができません。が、複雑な線を引く事ができれば、これはそれほど複雑ではないですが、丸を引くと、これは完全に分ける事ができます。これは 2 次元空間で複雑な線を引くというような形でやっていますが、先ほど申しましたように、ニューロンネットワークはどれほど複雑な関数であっても近似できるので、これを多次元空間でやるという事になります。医学のデータが 3 次元空間であっても、それを 100 次元に拡張してやり、その 100 次元空間の中で複雑な線を引くという事ができれば、心エコーなどの複雑なデータから病気を検出するというようなものが作れるという事になります。

この背景に基づいて 5 つの私の研究を紹介します。まず、心電図から心筋梗塞を検出するという AI について説明します。これは、最初に私が AI に着手した時に行った研究です。心筋梗塞は、心臓の周りで栄養を送っている血管が詰まってしまうという病気ですが、一度詰まってしまうと、心筋にダメージが蓄積していきます。血流を再開するまでは、この壊死というものが大きくなっていきます。心臓は再生しないので、心筋梗塞の治療で何が大事かというと、とにかく壊死をさせないとい

う事が重要になります。カテーテルを使ってやると、この詰まった血管を風船で一気に開けてやるという事ができるので、ここまでの時間を如何に速くして、壊死を減らすかという事が非常に重要になります。

そこで AI が活用できると考えたわけですが、心電図は 10 秒くらいで撮れます。胸が痛いという患者さんが救急外来に来たら、最初に撮る検査なので、心電図からそもそも心筋梗塞かどうかという事を見極めてスピードアップしてやれば、この心臓を救えるのではないかという事で最初に考えました。どのように学習するかという事を考えた時に、通常我々が見ている心電図は、ここに示している波形の画像で、紙に印刷されてくるので、画像として認識しているわけですが、画像で見ると、例えば波同士が被ってしまっていたり、印刷の品質が悪かったりなどという問題があり、情報として十分に実はきちんと残っていないかもしれない。心電図とは、心臓の電位情報を時間ごとに、2 ミリセカンドごとに記録した電位の時系列というものが本来の心電図のデータですので、これをそのまま学習してやろうという事で、本プロジェクトを始めました。

これが分かるとどのようなニューロンを使うべきかという事が分かってきます。通常のニューロンを全結合ニューロンと言いますが、入力に対して全部のニューロンに接合させるというような事を全組み合わせでやるという事が起こっています。これはかなり強力ではありますが、例えば最初のニューロンと 2 番目のニューロンというような、番号、順番のような事は特に分からなくて、時間とニューロンの対応は分からないというようになってしまうので、時系列を学習するには、実は適しません。

そこで、このリカレントニューラルネットワークを使います。これは展開した図になります。入力を 1 つのニューロンに入れて、他のニューロンには入れない。次は 4 秒目の入力を別なニューロンに入れるというような感じで入っていき、ニューロン同士が一方方向性の情報の繋がりを持っている事によって、自分が何番目の情報を学習しているのかが分かるというような仕組みです。これを使って学習させてやりました。これを使うと、自分が何番目か分かるという事は、時系列を学習できるという事になりますので、時系列データの扱いに優れているという事になります。

これは学習のコンセプトですが、ACS というものを心筋梗塞だと思って頂ければ良いと思います。急性冠症候群の事で、心筋梗塞だけではなく、心臓の血管が詰まる病気を全て網羅していますが、心電図を撮り、訓練された専門医による診断を AI で上回る事ができるかを試しました。

胸部不快感を訴えて、慶應病院の救急外来を受診した患者さんを対象にしました。48 時間以内に血流の再灌流(かんりゅう)療法をした患者さん、必要だった患者さんと不要だった患者さんをラベルにし、学習させました。そうすると大体 4 万人くらいの患者さんが、2012 年から 2018 年の 6 年間の間に慶應病院を胸痛、胸の胸部不快感という事で受診されていました。もちろん胸が痛いという事なので、全患者で心電図は撮られています、実はその中で再灌流が、いわゆる本当に心筋梗塞だった人は 400 人、1%くらいしかいません。これを夜間の救急の場でこの 1%の患者さんを見逃してはいけないが、全員カテーテルをやるわけにも行かないので、ここにかなり AI が生きる道があるという事がこのデータからも分かります。

AI モデルの学習ですが、入力、先ほど言った心電図の時系列データ約 4 万件を使って治療していれば 1、していなければ 0 という事で学習させてやり、こちらが結果になります。このテストコホートと書いてあるものは、AI は一度自分が見たデータに関してはズルをしてしまうというか、覚えてしまっているので、正解率がとても高くなってしまうという問題があります。ですので、事前に患者さんを学習用とテスト用というものに分けておき、これは一度も AI が見た事のない患者さんに関しての検査結果という事になります。これは、ROC 曲線と言いますが、縦軸が感度です。感度というのは、どのくらい見落とさないかという事を示すような所見で、本当に病気がある人のうち、何%を病気と判定できたかという事になります。もう一つが(1-特異度)という事で、特異度の逆数のようなものだと思って頂ければ良いです。特異度とはどのくらい正しいか、病気が無い人をどのくらい正しく病気が無いと言えたかという事を表す指標になります。これで曲線を描くと、この 1 に角に近ければ近いほど精度が高いという事になるわけですが、AUC0.88 と非常に高い精度を出す事ができました。一般的に大体 AUC0.7 くらいから比較的良いモデル、良い検査と言われている、0.8 であるとかかなり良く、0.9 だとかかなり完璧に近いというように捉えるのですが、0.88 なのでかなり高精度で捉えられるという事が分かります。実際に現場で使われている TIMI risk score というものがあるので、それと比較してみました。TIMI risk score というものは、感度が 97%くらい、特異度が 25%とされています。AI モデルで感度を 97%まで上げてやると、特異度が 40%くらいまで上がりますので、TIMI risk score と一般的に使われているものよりも更に高精度で分かるという事が示されました。ですので、AI モデルは心電図だけからで他の情報を一切使っていません。TIMI risk score は採血なども必要で、採血は結果が出るまで待ちますので、それと比べても、なおより高精度で検出できるので、これを使ってやると、先ほど言った治療前の時間を短縮でき、予後を改善できる可能性があるという事で、最初の AI としてこれを発表しました。

心電図からいろいろとできるという事が分かってきたので、ここでいろいろな病気をどのように検出してやるかという事を考えて、今度は ASD 検出です。ASD とは心房中隔欠損症の事です。これは大体、新生児の 0.07%にあると言われていますが、実は最多の先天性心疾患です。0.07%というと、低いように思うかもしれませんが、全人口の 0.07%にあるという事なので、人数としては結構な数があります。これは症状が乏しく、検診時に偶発的に診断されます。心電図異常があったり、聴診で異常な音が聞こえたりという事があると検出されるわけですが、これは稀で大体 7 割見落とされているという現状があります。多くの患者さんは、高齢までは見落とされていて、高齢で症状が出てきて、初めて見付かるという事になります。

出て来る症状は、例えばこの右の心房に血流が余計に入ってしまうと負荷がかかり、ここが大きくなってしまいますので心房細動というものが起こったり、ここにそもそも血栓ができて、ここを通り抜けて頭に飛んでしまったり脳梗塞になったりなど、そもそも負荷がかかり心不全になったりなど、肺に血液が行き過ぎて肺高血圧症になったりなど、このようなものは不可逆的なので、一度なってしまうともう戻る事がないので、事前に見つけて治療してやるという事が非常に重要な疾患です。どのようにやっているかというと、医師が心電図をスクリーニングして、検診で 40 歳以上の方は心電図を撮ります

し、学校検診で心電図を撮る大きな理由は 3 つあり、1 つは致死性の不整脈を発見する事。もう 1 つは心筋症の発見。先ほど話題に上がりました肥大型心筋症であったり、あとは拡張型心筋症だったり、心臓の遺伝的な病気を発見する事。そして最後がこの心房中隔欠損症を見付けたいという気持ちがあり、実施されています。ですが、7 割見落とされてしまうので、これでも 3 割しか見つける事ができないという事が今の大きなリミテーションです。どのようにしたら確実に見つかるかというと、ここに書いてある心エコー検査をやれば、大半見つける事ができるのですが、心エコー検査というのは、技師さんが患者さんに付き添って 1 人当たり 30 分から 1 時間くらいかかってしまうので、これを検診の場で全員にやるという事は現実的ではないので、どのようにこの心エコー検査をやる人を見付けるかというところで AI を使おうという事になりました。

こちらは研究デザインです。学習データは心電図データで先ほどと同じように電位データを使いました。検証データを事前に分けておき、学習データを使って、心房中隔欠損症がある人を 1、ない人を 0 という事で AI に学習させました。この時は、データとして Brigham and Women's Hospital と慶應義塾大学のデータを両方学習に使い、両方とも検証データを事前に分けておいています。また、これで内部の検証をしたのですが、施設を跨ぐと AI は大幅に精度が下がってしまうという事があります。それが何故起こるかという、これはその施設では使えるという意味ではなく、技師の癖や機械の癖のようなものを誤学習してしまう事によって起こります。

例えば、心房中隔欠損症のように、医師があまり情報を持っていないものであると起こりにくいですが、有名なものと例えば肺炎があるかという時には、肺炎があると思って、既に診断がついている人に 2 回目のレントゲン検査だったりすると、少し状況が良くないので、座って撮りましょうなどという事が起こり、座って撮ったレントゲンは、座って撮ったレントゲンなりに少し違いが出るので、そうすると座っていればこの人は肺炎というような誤った学習を AI がしてしまう可能性があります。そのような事をそれが全部の施設で共通しているという事は、通常考えにくいので、そのような癖のようなもので、どのくらい誤学習してしまっているのかという事は、外部データを使って検証すると分かります。ですので、今回は独協大学埼玉メディカルセンターという、学習に全く使わない外部データを使って検証を行いました。

こちらが結果です。慶應義塾大学で AUC0.9、Brigham で AUC0.88 という事で、学習に用いたコホートでは少なくともしっかり高精度で心房中隔欠損症が検出できるという事が分かりました。ちなみにバツ印が入っていると思いますが、これは人間が心電図を読み、ASD を検出したらどのようになるかという事を実際に専門の先生に読んでもらい、感度と特異度を同じ ROC 曲線上に打っています。AUC の曲線は、この人間の点よりもずっと上側を通っていますので、同じような特異度を達成しようと思うと、圧倒的に高い感度を達成できますし、同じような感度を達成しようと思うと圧倒的に高い特異度を達成できるという事がこの図から分かります。

外部検証コホートです。独協大学埼玉メディカルセンターは AUC0.85 という事で、非常に高精度を保っていて、少なくとも人間よりはとて優れているという事が分かります。この下がった原因です

が、慶應義塾大学は心房中隔欠損症を閉鎖している大学です。カテーテルで閉鎖できるのですが、全国から患者さんが紹介されてきて、診断がついている人を受け取って、閉じるという事をやっているのです。慶應義塾大学に来る人は結構穴の大きさが大きい人という事が多く、元々少し簡単であると思います。逆に独協大学埼玉メディカルセンターは完全な市中病院なので、自分のところで ASD を閉じたりなどはしていないですから、一般集団で検診で使うとどのような感じかという事をかなり表しているのです。検診で使ってもきちんと使えるという事が分かりました。

ここで AI のピットフォールについてお話しします。この辺りは結構有名なので、多分多くの方がご存じの事かと思いますが、パンダがテナガザルになってしまったりします。一番左の画像を人間が見るとパンダとしか言いようがないという事で、パンダと言うと思います。もちろん AI もこのまま見せるとパンダと言います。この真ん中にあるノイズのような画像自体、何の情報も無さそうで、これを足した一番右の画像は何か足されているが、人間の目から見ると何も変わっていない。これはパンダでしょうという感じですが、これを AI に見せると、99.3%でこれはテナガザルですと言うので、AI は人間であつたらあり得ないような、わけの分からない間違いをしてくる事があります。ですので、精度が例えば人間が 80%のところ AI は 90%になりますというと、10%は良くなるわけですが、間違えているものの中には絶対にあり得ないでしょうといった、ひどい間違いも往々にして含まれているという事です。

次の事例です。これは X 線から肺炎を検出する AI を作りましたという事で、非常に高精度で肺炎が検出できますと一時期センセーショナルに発表され、素晴らしいとなりました。しかし、蓋を開けてみると、これはどこを見ているかという事を見る GRAD-CAM という技術ですが、赤くなっているところを見ると、変なところを見えています。肺炎でしたら、この肺を見ていなければおかしいのですが、何故肩などを見ているのかというと、そこにこの技師さんが置くチップが置いてあります。このチップは何があるかという、前から見ているのか、後ろから見ているのかを表しています。患者さんが元気で、自分で歩いて検査室に行ける場合は、レントゲン室に行って後ろからレントゲンを撮ります。ですが、患者さんの状態が悪いと、患者さんも歩けないので、ベッド上で寝たまま前から写真を撮ります。そうすると、ベッド上で寝たまま前から撮っている人は調子が悪いので、肺炎であつたりする可能性が高いです。AI は前から撮ると肺炎だと学習してしまった。これは明らかに間違った学習をした例です。このような事が起こり得るので、AI がどのようなズルをし得るかというような事を常に考えなければいけなくて、そのためにはデータをよく理解していないと、医療の AI を作る事は難しいというところがあります。

最後の事例ですが、他施設に持ってくると、とても精度が落ちる事があります。これはメイヨー・クリニックが『Nature Medicine』に出した有名な論文で、心電図だけから心機能の低下を検出できますという内容です。これ自体は嘘ではないですが、ロシアの施設で試してみると、AUC が 0.93 であつたところが 0.82 と、0.1 落ちるという事は著しく落ちています。0.93 だとかなり完璧に近いのですが、0.82 ですと診断としては信じる事ができるというレベルではなくなってくるので、使えるものから使いにくいものに落ちるくらいの大きなインパクトがあるという事が分かります。この原因として

考えられる事は、技師の癖であったりという事があり得るので、必ず外部施設で検証するという事が重要になります。

AI はブラックボックスになってしまう。人間が教えなくても、データから自動で特徴を抽出するという、これは AI の優れた能力ですが、裏を返せば、抽出する特徴は人間がコントロールする事ができませんし、通常はどのような特徴を使っているか理解する事も難しいです。という事は、何でも使い得るという事になります。何でも使うという事は、例えば患者さんの名前、この名前は見た事があるというような事を覚えて、この名前の人は肺炎です、この名前の人は心筋梗塞ですというような事もする可能性がありますし、先ほどのレントゲンの前から、後ろからというような事もあります。あらゆるズルをし得るので、どのようなズルをしているかという事を人間がよく考えなければいけないが、それが何を使っているか分からないので、どのようなズルをし得るかという事も経験に基づいてやる以外にはやりようがなく、システムティックにやる方法が無いという事がもう 1 つのピットフォールです。

ですので、臨床で使える AI を作るために注意すべき事として、AI が適する問題に適用する。これはもちろんそうですが、対象とするデータをよく吟味し、特徴を理解し、AI がどのようなズルをし得るかという事をよく考えるという事が大事になります。ズルをしているかどうかという事を検証する一番簡単な方法は、外部データで検証する事だと思います。外部のデータでズルの仕方が完全に共通するという可能性はかなり低いので、特に国などを跨いでいるところは非常に良いです。日本で作った AI がアメリカでも使えるのであれば、AI の癖が完全に共通している事はほぼ起こり得ませんので、使えるのであろうと理解する事ができます。

これは今の研究テーマの一つになっていますが、可能ならば、AI が何を見ているかを明らかにしたいという事になります。という事で、今の特徴を踏まえて、今後、私がやってきた研究で、どのようにやってきたかという事をご説明します。ここで、心エコー、心電図から心アミロイドーシスを検出する AI について説明します。

心アミロイドーシスという病気では、最近新たな治療薬タファミディスというものが開発され、治療ができるようになっていますが、元々は治療が難しい病気でした。アミロイドというたんぱく質が心臓にくっついてしまって、そこで障害を起こし病気になるというようなもので、有名なところではアントニオ猪木がこの病気で、直近で亡くなっています。疑う事が難しく見逃されがちで、疑えば精度の高い検査ができ、治療もできるという事になりますが、これも心臓が不可逆的、心臓が再生しないという事が一つあり、1 回アミロイドがくっついて障害されてしまった心臓は戻らないので、重傷になってから見付けると、このタファミディスを使っても殆ど効きません。逆に心臓に殆ど影響が出ていないような状況で見付ける事ができれば、タファミディスを使うととても良く効いて、その後の進行をしつかり押さえられるという事になるので、早く見付けるという事が非常に大事です。ただ、疑う事はとても難しいです。そもそも心不全でかなり進行していても、心臓が少し厚くなっています。これは肥大型心筋症なのか、アミロイドなのかという事を区別する事は非常に難しいです。間違った診断がされていて、死ぬ間際になり、アミロイドーシスでしたとなる事が多いです。今の報告だと、診断がつく

までに何年もかかると言われていますし、医者も 1 人目だと診断がつかず、6 人目で診断がつくと
言われるくらい難しい病気です。ここの疑うという部分を AI にやらせてやる事ができれば、この病
気はかなり予後が良くなる可能性があるので、AI に適した疾患と考えました。

やり方ですが、心電図に関しては、先ほどと同じように時系列のデータを学習させるという事になり
ます。ただ、学習のさせ方で、今回は時系列をそのまま時系列データとして学習させましたが、誘
導軸、1 誘導、2 誘導、3 誘導と言い、12 誘導あるので、誘導軸をもう一方の軸として 2 次元デー
タとして学習させるという事を今回は使いました。この 2 次元畳み込みニューラルネットワークという
方法で、カーネルという画像のようなものを学習させます。カーネルはここに書いてある画像と似た
パターンに強く反応するという特徴があり、AI の場合は、この 3×3 くらいの小さいカーネルを何層
にも亘って組み合わせる事によって大きなものを学習するというような事をやっていきます。浅い層
では、まず縦線や横線のような非常に簡単なパターンを学習し、2 層目では今度はそれを組み合
わせて角を検出するというような感じで、だんだん複雑になっていって、最後の最後で非常に複雑
な心電図の波形を検出するというように、どんどん層を深くしていくと複雑なパターンを検出でき
るようになっていくという技術です。

ネットワーク構造ですが、この 2 次元畳み込みニューラルネットワークもいろいろな深さのものを組
み合わせて 1 つのモジュールにして、それを複数層重ねるといような、インセプションネットワーク
というようにものに近いような構造で、ただ心電図は少し特殊な形をしていて、12×2,500 のような通
常な画像ではあり得ないような形をしているので、独自のネットワークを使って学習させました。

患者集団は、各施設でカルテのデータを確認して、心臓の生検、実際に組織を取ってきて顕微鏡
で見たり、心臓 MRI やシンチなどで確実に心アミロイドーシスと診断された患者さんを抽出して来
ました。実は、心アミロイドーシスの患者さんは大体、電動障害が起こる。要は心臓の電気がきちん
と流れなくなるという事が起こるので、ペースメーカーが入る事が多いです。ですので、ここで AI が
ズルをし得る。ペースメーカーが入っている患者さんは、全員アミロイドーシスと言いかねないので、
ペースメーカーを入れた後の心電図を全部除外しました。逆側は心アミロイドーシスの無い集団か
ら年齢と性別をマッチさせた集団を対象群として抽出してきて、それを 0、1 という事で AI に学習さ
せました。この時は、Brigham and Women's Hospital に 1 万例あったので、このデータを学習
に使って、残り 2 つ、マサチューセッツ総合病院と UCSF のデータは検証に使うという形で行いま
した。

こちらが結果です。右側に ROC 曲線を示していますが、AUC は 0.91。施設によって少しずつば
らつきはあるものの、0.85 から 0.91 の間で、心電図のみからでかなり高精度で検出できるという事
が分かります。実際、こちらにどのような心電図を見ているのかという事を見ると、アミロイドーシスは
電位が下がってしまう事が多いのですが、そのような患者さんは非常に高精度で検出できていま
したし、逆にとても正常に近いような心電図でも AI がアミロイドーシスと言っている患者さんをよく見
ていくと、アミロイドーシスであるという事もあったので、人間でも見付けられないようなものを見付け

られるという事が分かりました。

これをこのまま使えるかという事です、心臓アミロイドーシスは全体の人口の約 0.5%で結構な数
がいるとは言え、それなりに稀です。そうすると AI に限らずですが、あらゆる検査で、検査の精度
がどれほど高くても、陽性的中率といって、実際にこの人は陽性ですと、検査で言われた人達の、
本当に何人が病気を持っているかというものは有病率に左右されて、有病率が低いと、とても下が
ってしまいます。0.5%などは全員にエコーをやるとしたら、0.5%しかももちろんいないわけですが、
この心電図でスクリーニングしてやると、それを 10 倍くらいにはする事ができます、しかし 5%なの
です。そうすると、100 人中 5 人しか本物がいないという事であると、少しまだ難しいかという事で、
もう一つ心エコーもその後にやり、心電図で陽性になった人に心エコーをし、それも AI でやるとど
うなるかという事をやりました。

ここで心エコーの動画を学習させるという手法を考えました。少なくとも当時、今でも結構難し
いところはあると思いますが、動画を学習させる手法というものが全く確立されていませんでした。最初
に出て来た心エコーの学習はフレームごとにばらばらにし、それを学習させるというような事をやっ
ていました。ですので、このフレームだけを見てこの病気ですかというような事をやらせていました。こ
れでも例えば心臓の大きさのような静的なものでしたら高精度で分かっていたし、アミロイドも心
電図と同程度、0.87 くらいの AUC では分かるという事が分かっていたんですが、これは難しいです。

これはマイケル・ジャクソンの写真ですが、この 1 フレームを見て、このマイケル・ジャクソンはどち
らに向かって進んでいるのでしょうか。我々がこれに答えられるかと言うと結構難しいと思います。AI に
フレームごとにばらばらにして学習させるという事は、このような事を聞いているわけで、これはでき
なくても仕方がない。これはフレームが沢山あったら動かしてみると、明らかにこれはムーブメン
ツで後ろに下がっているという事が分かります。動画として見せてやらないと、このダイナミックな情
報が失われてしまうので、エコーを動画として学習させる事は非常に重要だという事で、私は留学
して最初に動画を学習させるという事に着手しました。

どのようにやるかという事です、心エコーというものは、2 次元にプラス時間軸なら 3 次元の情報
です。それを 3 次元空間に構築して回してやるとこのように見えます。こうすると動きが 3 次元空間
の形として出て来るという事が分かります。模式図にするとこのような感じです。例えばこの弁の動
きを見てやると、弁の動きを 3 次元、時間軸方向にまとめてやると、ここに書いてあるような 3 次元
の形になり、この傾きが強ければ速く動いているという事です。傾きが緩やかであれば、ゆっくり動
いているというような感じで動きを形として検出できるという事になります。これができるという事にな
ると、これで 3 次元畳み込みニューラルネットワークを使うと、3 次元畳み込みニューラルネットワ
ークは、CT などで 3 次元の形を検出するために使ったネットワークですが、動画であっても 3 次元の
形として検出できるという事が分かります。

ただ、ここでまた動画だという事で難しくなる事は、心エコーはフレームレートがばらばらなのです。
50fps と 25fps、これは全く同じ心臓の心エコーを 50fps で撮った場合と 25fps で撮った場合を出

していますが、50fps だととても遅く見えて、25 だととても速く見えます。これをそのまま学習させてしまうと、AI は 25fps の方が速い、50 fps の方が遅いというような誤った学習をしてしまうので、これを全て 30fps に揃えました。学習は先ほどもあったこの 3 次元の畳み込みニューラルネットワークを使って学習をして行きました。アミロイドーシスの選び方ですが、基本的に同じようなやり方をしていますが、エコーは数が揃うので、逆側に今度はペースメーカーの患者さんを全部除くのではなく、ペースメーカーの入った患者さんをコントロール側、アミロイドのない側に寧ろ増やしてやる事により、アミロイドーシスはペースメーカーが入っているという誤った学習をさせないようにしました。あと、アミロイドーシスは心臓が厚くなりますが、厚くなる他の病気と区別させるために、HM、肥大型心筋症の方をあえてコントロールの方に入れる事によって区別させられるようにという学習をしました。施設数はこちらの方が少し多いですが、Brigham が学習コホートで、残りの 4 施設が全部外部検証コホートという形で使っています。

こちらが結果です。AUC が Brigham で 0.96、残りで 0.89 から 1、1 だと完璧ですが、かなり高精度で検出できるという事が分かりました。素晴らしい事は 1 施設で学習していても他施設できちんと使えるという事で、きちんとどこに持って行っても使えるという事が一つと、アミロイドーシスの検出精度が高いという事が分かりました。更に面白い事に他の分厚くなる疾患と区別できるかという事を試してみると、一番上が HCM、肥大型心筋症です。肥大型心筋症とアミロイドーシスを AUC0.96、0.90、0.87、0.91 と非常に高精度で区別できますし、高血圧、これは心臓が肥大してしまう高血圧の方であっても、それも非常に高精度で区別できます。最後は透析腎です。透析になると不要な物質が体に溜まってしまうので、それが心臓にくっついて分厚くなり、とてもアミロイドーシスによく似た所見が出てしまう事がありますが、それも区別できるという事が分かりました。更に診断がついた時のもので学習しているのですが、診断がつく 1 日前、30 日前、90 日前、180 日前、360 日前と、1 年前まで遡って、その時点で診断がつけられるかという事を見てみると、この 365 というものを見ると AUC0.89 位で、この時点で今診断をつけている人たちの 1 年前の時点でも AI に見せてやれば、実はアミロイドの診断がついたという事が分かるので、これで診断を早期化できるのではないかと期待しています。

最後に人間と比較しました。エコーの専門医の先生に読んでもらい、どの程度の精度でアミロイドは検出できるのかという事を見ていきます。感度として見ると約 60%。特異度は結構高いので人間は見落とすが、この人はアミロイドですと言ったらほぼ間違っていないという状態ですが、モデルを使ってやると、それでもなお特異度も高くなりますし、感度も高くなると。そもそも上を行っているのでモデル自体が優れていますし、カットオフを自由に決められるので、もう少し検出力を高めてやり、特異度は 75% で良いとなれば感度は 90% 位まで上げられるというような事で、殆ど見落とさないような使い方ができるという事になります。

つぎは多施設のデータを学習させるという連合学習という手法です。医療データは個人情報を含む・含む可能性がある為と言った方が良いのかも知れませんが、心電図自体を見て、この人は誰ですかと多分人間には分かりませんが、AI は分かっていたりするので、これは個人情報なのかどうか

というのが結構難しい議論のあるところですが、そのような恐れがあるので、データを提供することが非常に難しいです。これはアカデミア同士でも難しく、例えば慶應から Brigham にデータを持って行くという事はほぼ不可能で、患者さん全員から同意書を取らなければいけないという事になります。ニューラルネットワークによって構成される AI は、学習に大量のデータが必要なので、本来は多施設のデータを持ってくる必要がありますし、最初に申しましたように、技師の癖などを学習してしまう可能性があるため、可能なら多施設のデータを学習した方が良いのですが、この共有が難しいという事が 1 つの今の問題点です。

これを解消するために連合学習という手法を使って心電図や心エコーを学習させるという事を行いました。簡単に説明すると、施設 A、B、C と 3 つあるとして、それぞれにデータがあるとします。施設 A は施設 A のデータだけで AI を学習させる。B は B のデータだけ、C は C のデータだけを学習させます。でき上がった AI のデータを送らずに、重みだけを中央に送って、平均値を取るという事をします。このでき上がった平均値をもう一度この 3 施設に配り直して、次はその平均値から始めてもう一度自分のデータを学習させるという事を各施設で行い、また重みを送って平均化するという事を繰り返します。そうすると、最終的には一度もデータは自分の施設は出ていないですが、最後のこの AI は全施設のデータを学習したのと同じような効果が得られるという技術です。

これは心電図のデータを各施設で同じ方法で 2 次元にしているものですが、こうして見ると MGH、BWH、UCSF、慶應、似ているといえれば似ていますが、例えば慶應のこの部分などは完全に欠損していて、施設によって心電図であってもかなりばらつきがあるという事が分かります。これに基づいて AI に学習させた結果を見ていきます。

これは、心電図から肥大型心筋症を検出する AI を作っていますが、縦軸が学習に使ったコホート、横軸が検証に使ったコホートで、それぞれの数字が AUC を表しています。斜めが自分の施設のデータを学習し、自分の施設で検証した場合を表していますが大体、自分の施設で学習したデータは自分の施設で検証すると、かなり精度が良いのですが、例えば先ほどの慶應義塾大学のデータ、AUC が 0.93 と一番いいくらいで出ていますが、他の施設に持って行くと、AUC がとても落ちるという事が起きます。これは、多分癖を学習してしまっている誤学習が起こっているという事になります。これを防止するために 3 施設のデータを全部学習させるという事を先ほどの連合学習で行いました。そうするとどうなるかというと、これは、一番上に出ていた AUC の列は、自分の施設のデータを学習し、自分の施設で検証した時の AUC を表していますので、ここの斜めのところを表している事になりますが、これが連合学習したのですが、そもそも自分の施設を学習したものよりも更に精度が高くなるという事が分かります。

BWH、今回学習に用いていない連合学習はこの慶應、MGH、UCSF の 3 施設だけで行いましたが、外部コホートでやっても非常に高精度で肥大型心筋症を検出できるという事が分かりました。これは心エコーでも同様の事が分かります。ですので、多施設のデータを学習させるという事が非常に重要だという事が分かります。これをやるためには、連合学習というものは一つの手法として使

えるという事を示して、この結果は「Circulation」誌に発表しています。

AI の面白いところは、肥大型心筋症は心臓が肥大する病気なので、肥大しなければ診断が見つからないのですが、肥大が起こる前の心電図、心エコーであっても、今後肥大して行きます、肥大型心筋症ですという事を区別できるという事が分かりましたので、これも診断の早期化にとっても貢献しそうだという記述になります。

最後に人間と比較しました。こちらに書いてある事は、これは3人が心電図と心エコー、いろいろな組み合わせで診断をつけてもらった感度と、またこれは縦軸、プレジジョンというものが陽性的中率を表していますが、同じ陽性的中率を達成しようと思うと、感度はモデルの方が圧倒的に高いという事があります。

中身の分かるAIという事で、AIは最初はブラックボックスになってしまうという事で説明しましたが、これがどう解消できるかという話になります。例として心臓の機能の低下を心電図で分かるようなAIを紹介します。通常であれば心エコーをしなければ分からない左心機能の低下、駆出率の低下を心電図から検出できるという事をまずAIで普通に学習させると、AUC0.92、0.91、0.91、0.92と非常に高精度で分かるという事が分かります。

AIは何を見て心機能の低下をしているか判断をしているのかという事で、よく使われるものが、GRAD-CAMという技術です。先ほどレントゲンを前から撮っているのか、後ろから撮っているのかどうかを見せましたが、そこを光らせるという技術があります。これを見ると、波の高さを見ているのか、幅を見ているのかという事は分かりません。場所は分かります。例えば、これであるとQRSという波の形の部分を、場所を見ているというところまでは分かりますが、これがこの波の高さなのか、幅なのかというところは全く分からなくて、結局何を見ているのかという事は、この技術を使っても分からないという事が問題です。

最後に、このAIがどのように医療を変えるかという事ですが、内科医を置き換えるのではないかと良く言われていますが、私はその考えにはあまり賛同できません。と言うのは、医療はこの人にこれをやるのが絶対正しいという事はあまりありません。確率的にこのような事が起こる可能性が高いです、低いですというところまでは言えても、ではそれに対して治療しますかとなる事は誰かの判断です。それは通常患者さん自身が判断すべき事で、そこにしっかりとコミュニケーションでどのようにこのリスクを考えるべきなのかというような話をするのは医師の仕事です。そこは今でも一番重要な仕事ですし、変わらないであろうと思っています。ですので、今まで使っていたデータに更に追加してAIの結果も入って来る。1つのデータとして使うという事が今後の使われ方で、患者さんの診療をより良くするという事に使える新たな検査モダリティのような感じで捉えるのが良いかと個人的には思っています。

本日の結論です。AIは多次元データから効率的に情報を抽出する事ができるので、様々なピットフォールがありますが、かなり効果的な手法です。臨床現場での使われ方やデータを良く理解しな

ければ臨床上有用な AI が開発できないという事は、AI があらゆるズルをし得るからという事です、ここで AI の開発においても医師にも結構大きな役割があるのだらうと私は思っています。また、多施設データで検証しなければ、誤学習した AI に気付かないので、謝った判断に繋がる可能性がある、ここはかなり注意を要するところかと思っています。

今後の医療 AI の開発ですが、連合学習等を使って、多施設データを使っていく事が重要であり、また医療の知識と AI 開発に必要な情報工学の知識の両方を有した人材を育成し、これが臨床での活用やデータの性質を理解した上で AI 開発を行っていくという事を日本全体でやっていくという事が、少なくとも医療における AI 分野を発展させていくには重要だと思っています。また AI の中身を理解できる技術を発展させる。ブラックボックスを回避する事で、AI の活用範囲を広げるという事が重要かと思っています。以上です。ご清聴有難うございました。