

第 1 回

内分泌病理学における AI (artificial intelligence) (I)

臨床内分泌病理学会理事長

笹野公伸

はじめに

昨今の新聞、テレビ、インターネットニュースなどで "Dx", "AI" といったひと昔前にはほとんどお目にかからなかった用語が Covid19 と同じくらいの頻度で氾濫している。医学の領域も決して例外ではなく、極端な場合 Dx/AI に触れなければ時代に取り残される、あるいは今後の医学の進展を語れないような雰囲気も醸し出されている。しかし冷静にこれらの報道、特集又は審査を経ない商業誌の解説／総説などをよく読むと、中で使われている "Dx" や "AI", 特に後者に関してはその意味あるいは定義そのものがこれらの記事、解説間でかなり異なっている事に気づく。一方で医学領域での AI に関する情報は文字通り日進月歩を遂げており、時に劇的に変わる内容／進歩／情報にアンテナを十分張っておく事も必要である。

さてこの医学領域の AI に関し現時点で一番研究が進められているのは画像診断／解析の領域ではないかと思われる。特に artificial neural networks を事実上立ち上げた第一人者である Google Brain の Geoffrey Everest Hinton 教授が Google 他により開発中の deep learning image analysis software を用いる事で放射線科医は 5-10 年後には必要性がゼロになる事を、丁度深い谷の崖っぷちに追い込まれたコヨーテに例え説得力豊かに 2016 年に述べた事で一般からも大きな注目を集めた。さらに Hinton 教授は、このように AI が放射線診断の殆どを行えるようになるので、放射線科医のレジデントなどの卒後教育プログラムなどを廃止して他の診療科に医学部卒業生を振り分けるべきとも提唱し、多くのインパクトをマスコミを含む非医療関係の専門家に与えた。そして医学の領域外のマスコミ他からは放射線科医は歴史的価値しかない職種になるという、冷静に考えれば空想の戯言に過ぎない論旨が国内外で徘徊しだした。同じ画像を取り扱う病理診断の領域も決して例外ではない。例えばごく少数であるが東北大学の医学部

学生の中にも、普段からろくすっぽ講義にも出席しないばかりでなく成書／論文には目を通さずネットサーフィンなどで表層的に聞きかじっただけで、放射線／病理診断医の将来性はもうないのではと信じ、医学部卒業後誰かこれら将来性の無い領域に進む者などいるのですかね。。。。というような事を実際 SNS などで発信する者まで現れ、改めて上記のインパクトの大きさを認識させられた。

そこで本稿では病理診断学に関する AI の基本的な内容を現時点での最大公約数に相当する定義を述べ、本当に AI を病理組織／細胞診断に用いようとする際の利点と、種々の理由からか AI を推進する方々があまり触れられない問題／限界点に関しても考えてみたいかと思う。なお上記のようにこの領域はムーアの法則を凌駕するほどの日進月歩であり、ここで触れる事も会員の先生方が目に触れられる頃には一昔前になっている可能性もある事。この意味で筆者自身も 60 過ぎの手習いで勉強中である事。そして以下に述べるのは総説というよりは、筆者自身の経験などから導かれた個人的意見／コメントも多く含まれてる事。合わせて論旨的には 30 年以上に渡り病理／細胞診断に従事していた立場を無意識的に払拭出来ない事を予めここに断らせてもらう。

本稿では第一稿で、AI に関して基本的な事を述べ病理診断学に AI を用いる現況と今後の展開を問題／限界点も含めて触れる。そして第二稿では内分泌病理学の領域での診断、解析に AI がどのように用いられているのかを筆者の経験を含め具体例に触れる事で述べようかと考えている。

AI (artificial intelligence) / machine learning / deep learning

AI の始まりはいつか？

AI の定義、特に一体いつどのように生み出された

のか?という事に関しては色々と論議がある。実際種々の論文を見ても丁度黒澤明監督の羅生門のように各々の立場により真実とされる事柄も異なるように思われる。最大公約数あるいは一般的にはAIはcomputerの開発、発展と表裏一体と考えられる。特に単なる計算機能しかなかった自動計算機を1940年代に英国のAlan Turingが、ドイツ海軍の暗号装置であり人間の能力では解読不能とされていたENIGMA machineを介する暗号情報を解読出来る人工知能のレベルにまで引き上げた“Turing machine”が史上初のコンピューターとされる場合が多い。第二次大戦の初期にデーニッツ提督に率いられたDas Unterseeboot (U boat)がこのENIGMA machineを用い互いに連絡しあいWolfsrudeltaktik (wolf pack)とも呼ばれる狼の狩のように神出鬼没に現れ大西洋を渡る生命線であった英国船団の多くを海底に沈め大英帝国を降伏寸前にまで追い詰めた経緯を考えると、この評価も理解は出来る。しかしこれには第二次大戦の戦勝国であるいわゆるアングロサクソン国家の意向もかなり含まれている事は予想される事である。一方敗戦国であるドイツの技術者であるArthur Scherbiusが開発したENIGMA machineも今から見直して見ると、キーボードを介して伝わる機械的シグナルと電気シグナルを見事に組み合わせた上記のTuring machineにも匹敵する非常にレベルの高い電子装置である。本邦との関わりではこのENIGMA machineは当時同盟国であった日本にも伝えられ日本帝国海軍でも改良され用いられた。あまり知られてはいないが、この経験が戦後の本邦での電子技術の発展に大きく貢献したが、敗戦国の嵯峨かENIGMA machineの学問/技術的価値はほとんど無視されてきている。

このTuring machineはENIGMAを打ち破ったのだから何でも出来るはずと過剰な期待をされてしまい、1950年代の東西冷戦下の米国で膨大な予算がつぎ込まれAI研究としてソ連が次にどのような行動を起こすだろうかという予知に用いられた。しかしキューバ危機、ベルリン危機などに際して何の役にも立たないというお門違いの米軍の判断から米国政府レベルでのAI予算が一変に大幅削減されAIへの期待もかなりしぼんでしまった。しかしTuring machineが出来る事は十分はたしたと言うのが現在の評価である。

Symbolic artificial intelligence の誕生

しばらくして“AI”と位置付けられたのがいわゆる種々のデータを電子機器に入力する事で機器にある種の解答を出させるいわばsymbolic artificial intelligenceと呼ばれるプログラムである。このsymbolic AIは1970年代の終わりから1980年代にかけて隆盛をむかえる事となった。この中で有名なのがExpert systemであり、特にStanford大学で開発された簡単なinference engineを用い感染症の患者の病態をGPに相当する総合診療の医師が入力していくと、どの細菌に感染している、どの抗生剤が効果があるのかをその確率と共に示すMycinは大きな関心を集めた(1)。この背景として当時の米国で医師不足の領域で専門家の知識を蓄積して推論や問題解決を行おうとする電子機器の開発が連邦政府の膨大な予算をかけて行われた事情がある。中でも当時は感染症に対し次々と開発されてきたどの抗生物質を用いれば良いのかというところに多大なるエネルギーが割かれた。このMycinも感染症の専門家がいらない施設でも種々の臨床情報を入力する事で感染症専門家の判断との一致率は70%に達するとの報告がなされ、総合診療医でも感染症の専門家と同じ臨床診断と治療が出来ると一時期大きな注目を集めた。しかし実際専門家とMycinの間で選択された抗生剤の治療効果のいわばhead to headの比較検討試験は行われておらず、陰性コントロールの置き方、再現性などの点で今から見直すと片手落ちの欠陥データであった。

しかし現在のAIに対しての評価の状況とも共通するかもしれないが、未知のものよく知らない事に対しては批判がしにくい事から、上記の欠点/問題点はいわば見て見ぬ振りをされ、このsymbolic artificial intelligenceに相当するexpert systemは他の医療領域にも進出していく事になった。本邦でも1980年代にexpert systemに種々の情報を入力する事で医師よりも正確な臨床判断が可能になるという当時の“専門家”(今から見直すと偽専門家であったが。。。)のいわば御神託が行われ、この付度もあり当時の通産省も第5世代コンピュータとして血税から莫大な研究予算がつぎ込まれた。筆者も当時東北大学医学部の内科の講義で今後はexpert systemが発達する事で医師のいわゆるムンテラは不要になり、種々の検査データ他を電子機器に入力する事で臨床診断はつけられるようになるので、東北大学医学部卒業生たる者主に医学研究に従事する事が望まれるという大いな

る誤解ではあったが“善意”からくる“激励”を教授された記憶が今でも鮮明にある。しかし今から振り返ると十分予想されることではあるが、このMycinも結果的にはなかなか想定ほどうまくいかなかった。確かにMycinは実際の患者から採取した細菌のsmear, 培養のdataの所見からどのような菌と同定し、どの抗生剤を処方するのが患者にとり最善かの決定に必要な根拠(ガイドライン、文献他)を吸い上げ 当時としては最大限の配慮を重ねたプログラムであった。すなわちこのプログラム自体は十分評価すべき内容であったと考えられる。しかし肝心の患者の全身状態、細菌のデータ以外の臨床検査所見、画像診断所見、専門家のいままでの臨床経験などは一切入力しなかった為に、いわば木を見て森を見ずの状態であった。これらの理由から十分予想される事ではあるが、Mycinで規範された抗生剤が全く効果がなかった症例が相次ぎ一部では訴訟にも発展した。一方上記の内科医が将来不要になるとされたexpert systemであるINTERNIST-1も 対象とする患者の診断後の臨床経過はexpert systemが下した診断よりも、実際に患者を診療している内科医がしっかりと問診し、診断した診断名とより有意に一致するという報告が相次いで出され(2)、実際の診療現場では用いられる事はなくなった。しかしこのexpert systemなどのsymbolic AIは電子カルテの領域でその概念が広く用いられ歴史的価値は評価すべき内容とされている。

一方病理診断の領域でもこのexpert systemは導入された。当時オランダのFree University Hospitalに在籍されていた非常に優れた病理学者であったBaak教授らにより提唱された"Pathology Expert Consultation System"(PECS)と呼ばれるシステムであった(3)。これは例えば、非常に簡単な例を挙げると肺の気管支から採取された生検標本を病理医が見て、“肺”、“生検”というような項目を押すと、1項目目の質問として得られた標本の病理組織所見は細胞間の結合性があるかどうかという項目が画面に出てくる。結合性があるとなると、有りというボタンを押し、次に核の挫滅があるかどうかの質問が出てくる。そこでそれほど顕著ではないと言う項目に回答すると、次に病変の腫瘍細胞に異常角化があるかどうかと言う設問が出される。最終的にありと判断してボタンを押すと、90%の確率で該当する病変の病理組織診断は角化型扁平上皮癌というようなコメントが出てくるシステムであった。1990年代に東北

大学の故高橋徹名誉教授が開催された国際学会でBaak教授が仙台に来られて、このPECSを講演なされその後色々話をした事を筆者自身はBaak教授の時間厳守の厳格さと共に今でも覚えている。仙台でBaak教授はこれから10年で世界中すべての病理検査室でこのPECSが導入されると力強く述べられていたが、その後上記のMycin同様に病理診断のトレーニングをしっかりとする事でこのPECS以上の成果を効率よく挙げる事が出来る事が報告され、今ではこのPECSを知る者もほとんどいないのが現況ではないかと思われる。

AlphaGo と 鳩による乳腺の病理診断

上記のようにsymbolic AIを用いた病理診断への応用は1990年代後半になるとすっかり下火になり、AI診断もその専門家が曰く暗黒時代に一時的に陥った。しかし技術の進歩は絶えずなされており、21世紀になるとAIの病理診断に関連して2つの大きなbreak throughがあった。一つはよく知られているGoogle Deepmindによって開発された一種のComputer 囲碁プログラムであるAlphaGoである。詳細はここには述べないが、コンピュータが人知を凌駕する事は不可能であろうと考えられてきた分野の代表である囲碁の勝負の世界で、2015年から2016年にかけて人工知能が数多くの名人に勝利を取めた事は世間一般にも極めて大きなインパクトを与えた。そしてAlphaGoは世界的AIブームを呼び起こす契機となり、人が考え行おうすべての知的活動がAIによって置換されうるといふ大いなる幻影が徘徊する副作用も同時に生み出した。

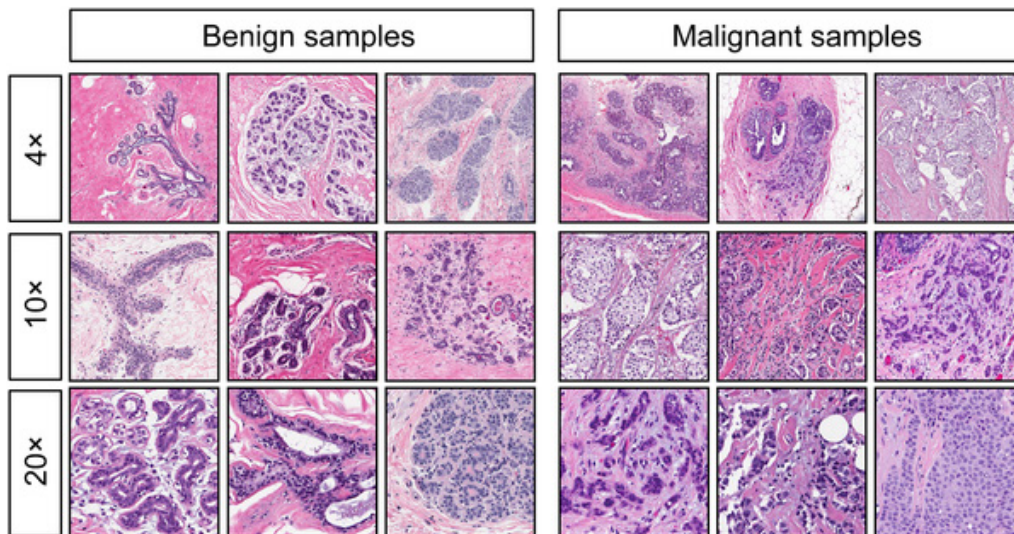


図1 文献4で鳩の学習／トレーニングに最初に見せた乳腺組織／疾患の病理組織所見。 Benign samplesはこのヘマトキシリン-エオジン染色の組織像で見てもよく発達して fibroadenoma や乳管内過形成も含まれておらず、基本的には正常乳腺といわゆる fibrocystic disease に相当する組織所見である。 Malignant samples とされている悪性症例の病理組織所見も核異型度も高く細胞密度も亢進しており、腫瘍細胞の集塊の辺縁も不規則であり、いわば医学部の病理学の学生実習標本としても適している乳癌の病理組織所見である。

もう一つが鳩を訓練して乳腺疾患の放射線画像診断と病理組織診断を行わせた米国の University of California Davis からの研究報告である(4)。詳細はこの興味ある論文を拝見してもらいたいですが、図1に示すような乳腺の良性病変、悪性病変を鳩に見せて食事条件などで正診を学習させた。その結果同じ画像のセットで乳腺病変の良悪性を15日間の学習で鳩は80%を超える正診率で診断するなど適切に病変を認識したばかりか、画像自体をモノクロにして色の情報を喪失させたり、組織画像を回転させて鳩の空間認識能を混乱させてもほぼ同じ正診率を得られたという研究報告である。更にこの研究グループは類似した乳腺の良性／悪性疾患の病理組織画像を見せ良悪性の判断を鳩にさせてところ同様の正診率を得られたという驚くべき内容であった。興味深い事にマンモグラフィーなどの放射線画像診断の面では鳩の認識では病理組織画像よりも低い結果が得られている。

この論文の研究計画は非常によく練られていて、その目的も既に実証済みである基底核などの神経経路を含めて視覚を介する記憶力が動物の中でも非常に優れている鳩を実験台として、ヒトの画像認識の基礎を検討する事でよりよい medical imaging hardware の開発に結びつけようとしたものであった。しかしいつものパターンでマスコミがこの研究成果に飛びつき曲解し、もし鳩

に病理組織診断が正確に出来るのであれば、これだけAIが発達したのだからAIを用いれば完全に出来るはずという論旨が徘徊した。このような論旨に対して考慮しなければならないこの研究報告の問題点としては以下の3点が挙げられる。第一点はこの実験で用いた鳩は神社などにいる普通の鳩ではなく、非常に特殊な訓練を受けた鳩である事が挙げられる。第二点としては図1に示すように、鳩に見せて学習させた良性／悪性の乳腺疾患の病理組織像は非常に典型的な所見である。具体的には東北大学医学部3年生の病理学の試験で再々試験などに残ったごく少数の学生を、不合格のレベルの科目が病理学だけなので、一科目による留年を回避する為に何とか合格させようと出題するいわば救済レベルの病理組織像の問題とほとんど変わらないレベルであるという事もここで強調したい。第三点としてはどれだけ学習／トレーニングをしてもこれらの鳩が良悪性の鑑別が出来なかった病理組織所見があった事である。これらの所見を図2に示す。この論文でも骨子は報告されているが、筆者が見た限りではこれらの症例の画像に共通する項目としては病変の細胞密度が少なく、間質が豊富で正常の乳腺組織の乳管、小葉などが保持されている乳癌組織であった。すなわちパターン認識のみではこれらの病変は良性病変あるいは正常乳腺組織との鑑別が極めて困難であったと言う事が浮かび上がる。一方でこ

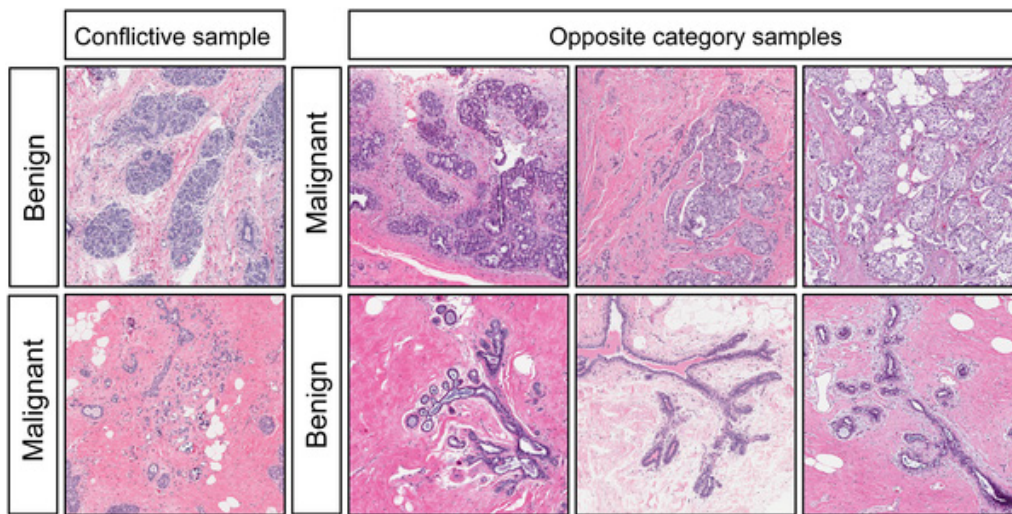


図2 文献4で鳩が何度学習しても良悪性の鑑別が困難であった乳腺疾患の病理組織所見。 Malignant とされている乳癌症例も乳管/小葉の基本構築はある程度保持されていて、個々の腫瘍細胞における核異型度も決して高くはない。また間質も腫瘍細胞の浸潤による desmoplasia などの所見も認められず一見正常乳腺組織の間質と変わらない組織所見を呈している。

の事は鳩ばかりでなく今後の AI を用いた病理組織所見の解析に重要な知見を提供するものとも考えられる。

AI/machine or deep learning と病理診断

AI (artificial intelligence) という用語が世の中に現在氾濫しているが、マスコミ関係などでは AI=deep learning (深層学習) とこの両者をほぼ同義語として用いている傾向もあり用語の定義を明確に正確に理解しておく事が医療関係者も要求される。

さて AI は前述のいわゆる symbolic AI を除くと machine learning (機械学習) とその中に事実上含まれる deep learning に大別される。そしてこれだけ AI という用語が使われると、この両者を適切に鑑別する事が重要だが現実的にはかなり難しい。その理由の一つとして、machine learning よりも deep learning がより高尚で困難そうな印象があるという事からか、実際に行なっているのは前者そのものであるにも関わらず deep learning を用いた解析とその結果などという論旨が病理に関連する研究報告でも多く認められる背景が挙げられる。すなわちこのような用語の使い方でもいわゆるハゲタカジャーナル以外の peer review system、editor の判断をすり抜けてしまい “published study in peer reviewed journals” となってしまう研究報告が少なからずある事実である。

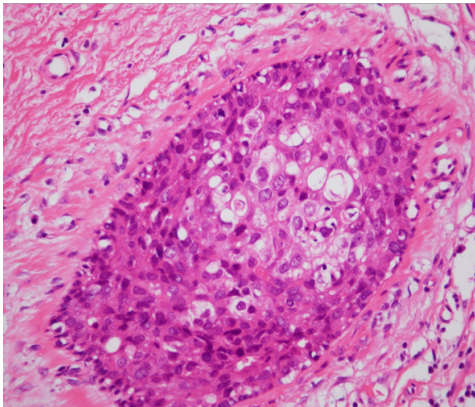
さて実際の病理組織診断を例にとってもう一度振り返ってみると、上述の病理関係の PECS を用いる

と、図3の乳腺腫瘍の CNB (core needle biopsy) 標本を病理診断医が最初に顕微鏡あるいは画面で見たあとで、computer が例えば核異型はあるか否か?、あればどの程度であるのか? など yes or no あるいは multiple choice の形式で病理診断医に問いかけてくる。病理医は自分で判断した内容を入力し、例えばこの図3の A,B であれば双方共に顕著な核異型が見られるという解答を入れて PECS の computer program が 90% 以上の確率で neoplastic な病変であると言う回答を出す。更に間質浸潤があるかどうかを computer program が病理医に問いかけ、A では明瞭ではなく B では間質の未熟な繊維芽細胞が多く認められる desmoplasia、炎症性細胞浸潤などから 間質浸潤 (stromal invasion) ありと入力する。そうすると program が A は high grade DCIS (ductal carcinoma in situ) の可能性が 98%, B は同じ high grade であるが IDC (invasive ductal carcinoma) の可能性が 99% と算出する。これを見て分かる通り、普段顕微鏡を見てしっかりとした卒後教育を受けた病理医であれば瞬時に判断する項目を入力という手間をかけて同様の病理診断に達するのが PECS を含む symbolic AI であり、何故この program が誰も使われなくなるようになったのかがよく理解出来ると思われる。

Machine learning

これに対して 狭義の machine learning、すなわち supervised learning は図3を例にとると病理医が各々

A



B

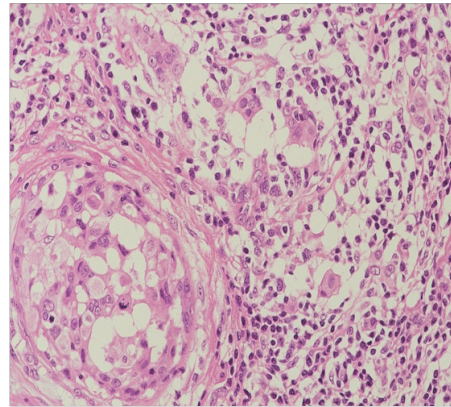


図3 乳腺腫瘍のCNB (core needle biopsy) の病理組織所見 (ヘマトキシリン- エオジン染色)

の症例の特徴を入力する必要は一切なく、乳腺疾患の病理組織標本の画像所見を取り込めば上記のA,Bの病理組織診断にまで達する事が出来る computer program である。このように記載すると如何にも自動販売機の中にお札を入れるように病理標本を差し込む事で病理組織診断がすぐに出るように報道するマスコミもあるが 後述するようにこのシステム自体にも数多くの準備、問題点がある。一番注意すべき点としては、この program に、核異型とは何を表すのか？、核異型度が高い、低いとはどのような事か？ 間質浸潤とはどのような場合にそう捉えるのか？など数多くの画像そのものをその解釈と共に program 入力させて学習/training させなければならない事である。当然の事であるがこの入力して学習させるデータ (training data) の種類と量が多ければ多いほど病理組織診断はより正確になる。逆に言うと training data の質と量によって最終的な病理診断の正確さが規範される。しかし細胞の染色動態の色調、核や病変の大きさなどの連続数値として示される因子をプログラムの中に取り入れるとなると膨大なデータが必要になる。この事は現時点では大きさが膨大な精緻な画像を圧縮しないで取り込まなければ、診断精度もかなり低下する事も示している。また実際この machine learning であっても病理診断医が日常行なっているプロセスをそのまま行なっているプログラムそのものであると言う事も十分認識しておく必要がある。しかし同時にこの supervised/machine learning は deep learning とは異なり病理診断に至る過程まではけして black box ではない。すなわち何故最終的な診断に至ったのかと言う経過、背景はある程度示す事が出来る。このような事か

ら現在発表/報告されている病理組織所見の画像解析、“自動診断”は実は deep learning ではなくその大部分が machine/supervised learning である。しかし後述する問題点を考えると現時点で病理組織診断にこの machine/supervised learning を取り入れるのは研究/学術的価値は確かに非常に大きいと思われるが、日常診療で全ての症例の病理診断に用いるのは非現実的と判断される。

Deep learning

Deep learning は以前は中枢神経の機能を computer program が行うと言う事から “artificial neural networks” とも呼ばれていたが、重要なのはあくまでも machine learning の一つであると言う事実である。上述の machine learning と異なる点としては、例えば病理組織標本上で細胞分裂とはこのような形態学的所見だとか、核異型度が高いとはこのような事だと言うような training を全く必要としない事である。すなわち形態学的所見から共通する特徴、例えば細胞の輪郭がどのくらいスムーズなのか、でこぼこしているのかなど病理医が通常表現するのが困難ないわば hidden layer に相当する特徴を computer が認識しそれを分類していく事である。この deep learning は microarray などで用いられる hierarchy clustering などが代表例としてよく挙げられるが、基本的には上記のような多くの hidden layer に基づいて解析を進める事から現時点では何故そのような病理組織診断に至ったのかと言うプロセスは不明のまま、すなわち blackbox であるところが実際の病理組織診断に用いるに際しては問題点となる。一方この deep

learning は病変の病理組織診断を得るという目的を入力すると supervised learning にはなるが、どのような指標を用いるのかなど細かい指示は上記のようには入力しない事から unsupervised learning と位置付けられる。

さてこの deep learning と machine/supervised learning を具体例でどちらに相当するのかをしてみる。上記の鳩による乳腺疾患の病理組織診断は computer と鳩の差異はあるものの前者である。AlphaGo は前者と位置付ける論調が多いが、冷静に考えると囲碁のルールや過去の対戦を training として学習させていて、実際にはありえない事であるかもしれないが、少しでも囲碁のルールそのものが変わったりすると対応は困難になるので後者と考えるのが妥当であろう。一方この machine/supervised learning が deep learning に自動的に転換する、あるいは進化すると言う事は現実的にはあり得ない。しかし医学領域ではない事から適切な例ではないかもしれないが、“2001年宇宙の旅”(A space odyssey)の一種の主演である人工知能 HAL にその事象は示されている。興味がある方々は SF の第一人者であるアーサー・クラークのアイデアでスタンリー・キューブリックが監督し 1968 年に制作され 筆者の経験から何度見ても難解で娯楽作品とは程遠いが“モノリス”など今でも考えさせられる内容を含むこの映画を一度ご覧になると良いかもしれない。何れにしても deep learning 特に unsupervised learning を用いた解析により病理組織/細

胞の形態学的検索/診断は今後新たな展開を遂げる事は間違いない。しかし後述するように machine learning 以上に日常診療/病理細胞診断に応用するには費用、手間などの点で大きな課題があり、診断に至る過程が blackbox である事に起因する法的/社会的な問題も加わり少なくとも近未来的には研究/学術目的の使用が主体であり、診療レベルでは補助的な位置付けになると言うのが現時点での順当な結論であると考えられる。

AI/machine learning を病理組織細胞診断/解析に用いる際の問題点

対象とする標本/検体の多様性/artifacts

現在の病理細胞診断/解析の主流である machine/supervised learning では上述のように如何に多くの情報量を training set として computer program に学習させられるのか? という点が非常に重要である。AI の応用が盛んに議論されているもう一つの診療科である放射線画像診断では、放射線画像の情報量は病理組織所見より絶対的に少なく、更に胸部 X 線、CT 撮影などで被験者/患者の体動下で得た画像しかないと言う事例はあまりない。しかし病理組織/細胞検体の場合、図 4 に示すように挫滅、乾燥など採取時あるいは標本作成課程での artifacts が認められる事例が決して少なくない。加えて気管支生検などでは、診断に支障がある artifacts が認

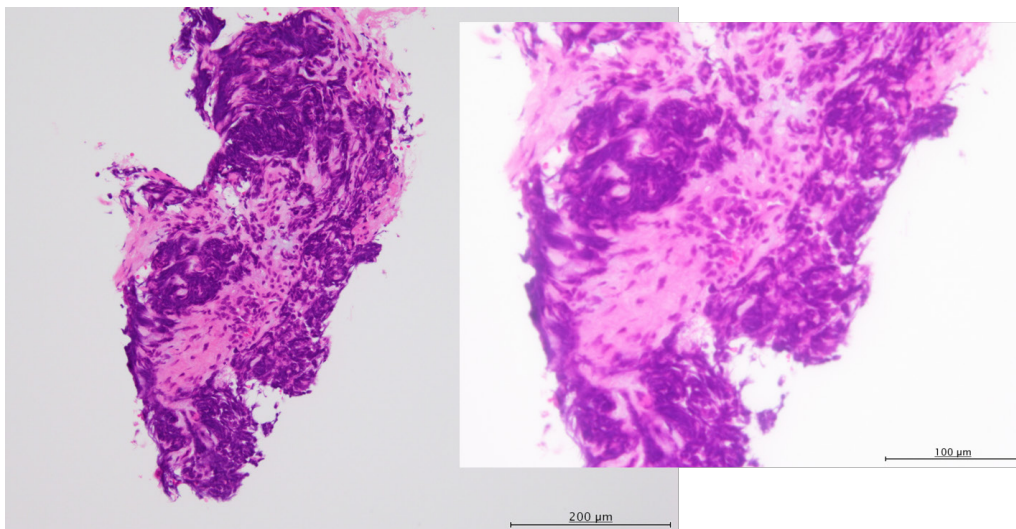


図4 transbronchial biopsy で採取した気管支原発の小細胞癌 (small cell carcinoma) の病理組織所見 (ヘマトキシリン-エオジン染色)。Neuroendocrine carcinoma の一つである小細胞癌は核の脆弱性は nuclear fragility, DNA stain として良く知られているが、採取時の種々の条件によってその程度は大きく影響される。このような多彩は artifacts の所見を computer program に予め training させる事は事実上不可能である。

められるからと言って該当する全症例で再検を求める事も現実的でない。すなわち少々 artifacts がある検体でも出来るだけ正確な病理組織/細胞診断を臨床側に示す事態も日常診療では決して例外ではない。更にこれらの artifacts は極めて多岐に渡っており、これらを全体的に computer program に training する事は事実上不可能である。

更に近年では免疫組織化学, in situ hybridization が病理組織診断に重要な役割を果たしてきている。これらの解析では preanalytical factor, 特に採取時の状態、固定までの時間、固定液の種類、固定時間などが結果に大きな影響を与える。特に近年タンパク質の抗原性保持に良好であると提唱され多くの施設で用いられる事が多い緩衝/中性ホルマリン固定液は、注意して標本作成を進めないと固定不良に陥りやすい。図 5-1,2 にリンパ球豊富型ホジキンリンパ腫のリンパ腺標本の緩衝ホルマリン固定不良の例を挙げるが、中心部に固定不良あるいは固定されていない部位が見られ、病理組織診断の要となる RS 細胞の形態所見も大きく異なっているのが分かる。更に近年ではホジキンリンパ腫の治療では従来からの ABVD 療法 (アドリアマイシン、ブレオマイシン、ビンブラスチン、ダカルバジン) に加えて CD30 に対しての特異的標的治療であるブレンツキシマブドチンが初回治療から用いられるようになり、病理組織検体での CD30 発現の有無が治療適応の決定に際しても非常に重要になる。この CD 30 発現の有無の検討に際しては当然固定条件が悪くなると偽陰性が多くなり、発現強度も固定条件に大きく左右される。病理医の診断ではこのような固定標本の微妙な固定ムラを考慮して CD30 の免疫組織化学の判定が経験次第では可能であるが、machine/supervised learning では温度によっても固定状態が異なる検体の preanalytical factors を全て training あるいは学習させる事は事実上不可能であり、病理組織診断への AI の応用に際し大きな限界点としてあげられる。

COST、費用、コスト

用いる program が machine/supervised learning であれ、deep learning であれ病理組織/細胞所見を AI で解析するには該当する標本をバーチャル顕微鏡でデジタル画像に取り込まなければならない。このデジタル画

像に取り込む事は、東北大学病理で大震災の際に生じたようにガラス標本のほとんどが破損したり、標本が退色したりしないので良い事尽くめのようにも思えるかもしれない。しかし日常診療の観点からは現時点では費用と手間の関係で問題点が山積みである。この中でも一番はコストの問題が挙げられる。2021年の USCAP (Unites States and Canadian Academy of Pathology) の年次集会で David McClintock 教授が述べていたデータを以下に記載すると、年間 5-6000 検体の病理検査室/病理診断科で仮定すると米国の場合で初期投資として年間約平均して 1,645,000 US ドル、その後の維持費で 5 年間で同様に平均して 3,625,000 US ドルかかるかと言う試算を出されていた。この初期投資費用としては 3 台のバーチャル顕微鏡とスキャナー、用いる AI program のライセンス代、患者の病理組織診断に用いる事からいわゆる medical grade の病理組織/細胞画像をしっかりと描出可能なモニターなどにそれなりの費用がかかるのはある程度予想される事であった。しかし病理組織所見をいわば保管しておくサーバーの費用が思いの外高いのには拝聴させていただき驚かされた。サーバー費用の内訳は膨大な病理画像の storage にかかる費用がほとんどであり平均して年間 150-200 TB 以上ないと円滑な診断業務が出来ないようである。更に驚かされたのはサーバー、computer などの冷却代を含む電気代である。これはビットコインなどの仮想通貨でも指摘されていた事であるが、機器が発生する熱は如何に回路他が進歩しても抑えられない事を再認識させられた。一方で米国の電気代は本邦よりもかなり低く抑えられている事を考慮すると本邦ではより高い維持費が予想される。更に論文他で評価済みの machine/supervised learning に用いられる program の多くが米国発である事を考慮すると本邦の癌拠点病院クラスの病理検査室/病理診断科で初期投資として初年度 2 億円、導入後の 5 年間の総計で約 4 から 5 億円費やされる事が現時点では想定される。しかも AI 診断を行なっても行わなくても病理標本の作成料、診断料、判断料、加算料として病院に入る収入は変わらないのが現況である。そして病理組織/細胞診断を積極的に AI で行なったとしてもその性質上患者の集院効果にはほとんど期待する事は出来なくむしろ法的なリスクを犯す危険性も否めない。このような現況を考えると、本邦での病理組織/細胞診断の診療報酬体系にコペルニクス的な変革が生じない限り少なくとも AI

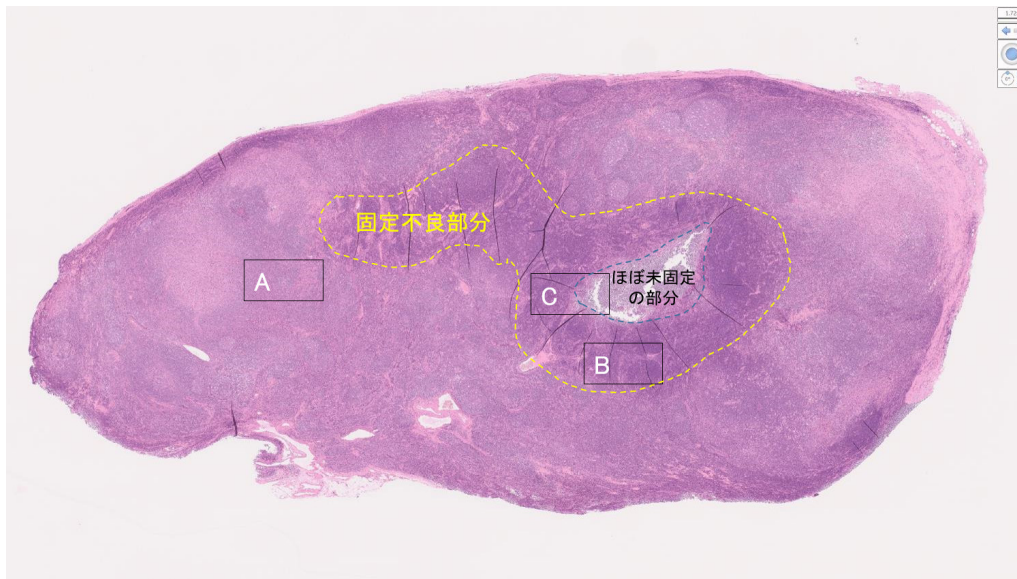


図5-1 リンパ球豊富型ホジキンリンパ腫のリンパ腺標本の緩衝ホルマリン固定不良の一例の病理組織所見（ヘマトキシリン-エオジン染色） 固定が辺縁部では十分されているが、中心部では黄色い線で囲まれた領域が固定不良部位であり、中心部にはほぼ未固定の部位が認められる。十分固定された部位（A）、固定不良部位（B）、ほぼ固定されていない部位（C）の高倍率像を4-2で示す。東北大学病院造血器病理学共同研究部門 一迫教授 御恵与

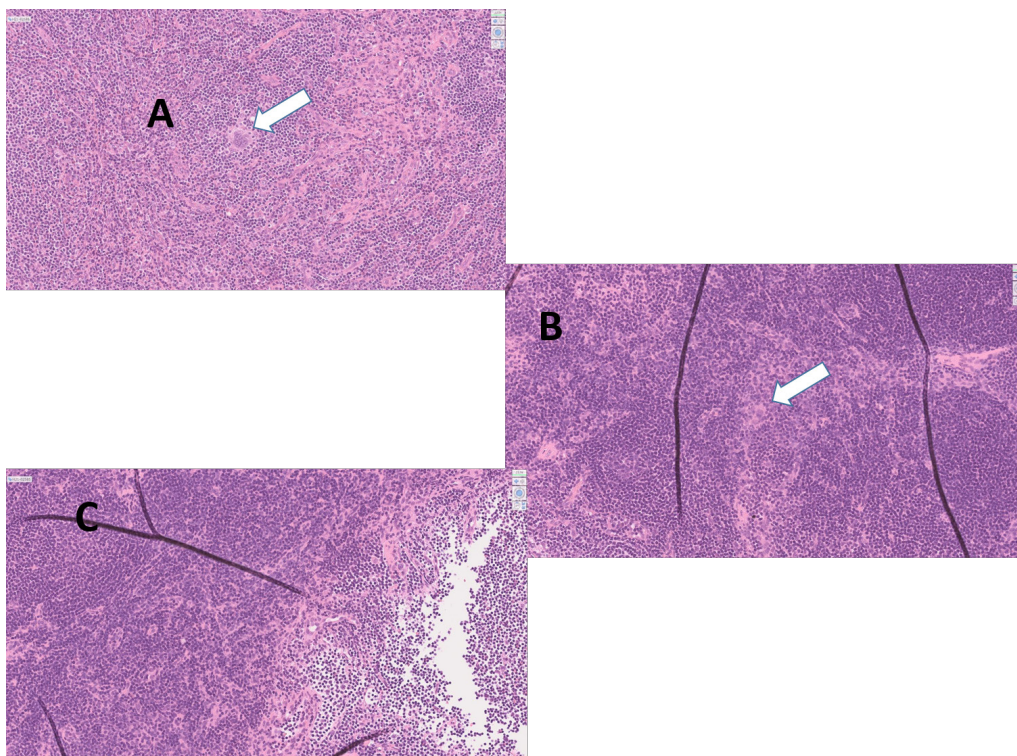


図5-2 リンパ球豊富型ホジキンリンパ腫のリンパ腺標本の緩衝ホルマリン固定不良の一例の病理組織所見（ヘマトキシリン-エオジン染色） 標本の固定条件によりリンパ球、ホジキンリンパ腫の診断で極めて重要になるRS細胞の形態学的所見も大きく異なってくるのが分かる。

の病理検査室／病理診断科での日常診療への導入は夢物語の更なる夢であると考えざるを得ない。しかし同時に上記のような事をほとんど考えなくても良いお隣の中国ではこの領域は日常診療への応用を含めて日々目覚ましく進んでおり、近未来を含む今後の事を考慮すると病理組織／細胞診断の研究／学術領域での面だけでも進めていく事が強く望まれる。

今後に向けて

病理組織／細胞診断における AI を用いた解析／診断は現時点ではいわば発展過程にあり少なからぬ問題点が上述のようにある。しかし今後技術の発達と共にいわゆる breakthrough が生じてくる可能性も十分想定される。このような事態に備え、今から本邦の病理検査室／病理診断科で行える事を、私見ではあるが以下の2つにまとめる。

1. 病理組織／細胞標本のデジタル化
2. 標本の固定、作成などの preanalytical factors の均てん化

上記の2点共に現実的には費用も手間もかかり、いわゆる机上の空論あるいは "academic exercise" という批判もある事は重々承知しているが、この2点につき少し最後に触れてみたい。

1. 上述のように machine learning であれ deep learning であれ病理組織／細胞所見を全てデジタル化する事が病理組織／細胞診断に AI を応用するに際しては必要である。すなわち病理組織／細胞診断あるいは解析には virtual 顕微鏡を用いた画像の取り込みが絶対的な前提条件となる。しかしこれに関しても前述の機器、消耗品、サーバーの費用の問題以外にも 現状では以下のような問題点がある。

(1) バーチャル顕微鏡で取り込んだ画像の互換性

バーチャル顕微鏡で取り込んだ画像所見を実際にモニターで観察するあるいは machine 及び deep learning を含む解析を進める為にはいわゆる viewer と呼ばれるインターフェースが必要である。しかし現時点ではバーチャル顕微鏡のメーカー、プラットフォームによりインターフェースが異なる。実際我々のところでは3つの異なるメーカーのバーチャル顕微鏡を用いているが、例

えば A 社で取り込んだ画像は A 社のビューワーでなければ実際に観察／検討する事は出来ない。更にいくつかのメーカーでは Mac では全く画像を観察する事が出来ない。この事はビデオデッキの黎明期に同じビデオテープであっても VHS と beta では互換性が全くなかったと同じ状況と考えてもらうと理解しやすいのではないかと考える。更に他の施設で取り込んだ画像を同じビューワーを用いなければ一切観察する事が出来ないという、AI/machine-deep learning を広く普及させるには致命的な欠点が生じている。この事に対しては現在 Part 2 で触れるが入力された異なる画像から新しい擬似データを生成する GAN (Generative Adversarial Network) がデータ不足を補足する手段として注目され幾つかの研究報告も病理組織所見の解析では報告されだしてきている。しかしこの GAN は基本的にいわば fake / 偽画像を作成して解析するアプローチである。この事から Journal of Steroid Biochemistry and Molecular Biology や Endocrine Review の editor の一人として blots や蛍光画像の manipulation を如何にして見破るかにかなりのエネルギーを費やしてきた筆者の経験からは 個人的にはこの GAN は邪道としか考えられず到底病理組織診断／解析の一環として受け入れられるものではない。しかしこの GAN を用いた解析を使用した病理関連の研究報告は日々増加しており、"嘘も方便" が常態化するのも時代の流れなのかもしれない。

現在どのバーチャル顕微鏡で取り込んだ画像でも観察あるいは AI を含む画像解析が可能になる free の software の開発が進んでいるとも聞くが、ここが解決されない限り病理組織／細胞診断の AI/Dx 化はまさに絵空事に過ぎない。

(2) モニターのグレード

前述したように米国ではいわゆる digital pathology の診断には medical grade のモニターが必須と位置付けられている。筆者の経験でも大まかには問題ないが、核の微細な構造や免疫組織化学の弱陽性、あるいは細胞膜に全周性に陽性所見が認められるか否かなどはかなりモニターの性能に左右される。現時点では medical grade とされるモニターは高価なものが多いが、この点は他の電子機器の今までの性能の向上と価格の推移を考慮すると個人的には上記の (1) の問題点と比較すると楽観視している。

2. 病理組織／細胞診断への AI 特に machine learning を応用する際にはどれだけ多くの標本の情報量を学習させるのかということが精度をあげる為には大きなポイントとなる。この点で標本毎の固定条件が異なったり、病理組織標本作成に施設あるいは検査室毎にばらつきがあると機械学習の対象となる病理組織／細胞標本の study set はほぼ永遠に出来ない。

上記のように病理組織／細胞診断一解析のプログラムの多くは現時点では米国又は中国発である。しかし欧米、中国、アジア諸国などからの数多くの病理組織標本をコンサルトあるいはセカンドオピニオンとして診断に関わってきた筆者の個人的経験から、機械／自動化が進んできた現在でも施設間の差異はある程度あるものの本邦での病理組織標本が尤も artifacts が少なく観察しやすいと思われる。少なくとも筆者がレジデントとして病理を習得した米国よりは病理標本作成に従事する臨床検査技師のレベルは間違いなく高いと自信を持って言える。しかし標本の乾燥、固定不良、挫滅など病理検査室レベルでの努力ではどうしようもない preanalytical factors は本邦と諸外国ではあまり差がない印象を持つ。

このように本邦の病理組織標本作成に従事する臨床検査技師の技術レベルは世界でも白眉であるが、今後の AI/ デジタル化を考慮すると以下の 2 点が現在出来る事ではないかとも思われる。しかしこれも上記の 1 同様にそれなりの費用が当然かかる訳で、これを実行するかどうかは病院／施設管理者が、中央診療部門の要となる病理診断科／病理検査室をどう位置づけるか、あるいはどう考えているのかの“リトマス紙”、“炭鉱のカナリア”となっている。

(1) 病理検査室の ISO, CAP などの第三者認定の取得

これを得たからと言って preanalytical errors が全くなくなるという訳ではないが、改善する事は間違いないと思われる。

(2) 病理組織標本作成、染色過程の可能な限りの自動化

上述のように現時点で平均すると本邦の病理検査室の標本作成レベルは非常に高い事は間違いない。一方で病理組織標本作成は地味であるが本人の経験により大きく左右される事も事実である。すなわち本邦ではベテランの力量によって支えられている側面が特に強く、現在の若い臨床検査技師を見ていると正直 10 - 20 年

後が不安と感じられる事は否めない。この意味で初期投資はかかるものの、免疫組織化学を含む染色過程の自動化を可能にする機器の導入が諸外国同様に必要とされるのではないかと思われる。

【文献】

1. V L. Yu, Le M. Fagan, M, Sharon et al., Antimicrobial Selection by a Computer
A Blinded Evaluation by Infectious Diseases Experts
JAMA. 1979;242(12):1279-1282
2. R A Miller, H E Pople Jr, J D Myers Internist-1, an experimental computer-based diagnostic consultant for general internal medicine N Engl J Med 1982 307(8):468-76.
3. J P Baak , P H Kurver Development and use of a rule-based pathology expert consultation system Anal Quant Cytol Histol 1988 10(3):214-218.
4. R M Levenson , E A Krupinski , V M Navarro , E A Wasserman Pigeons (Columba livia) as Trainable Observers of Pathology and Radiology Breast Cancer Images PloS One 2015 Nov 18;10(11):e0141357. doi: 10.1371/journal.pone.0141357