

KONRAD WOJDAN*, MARCIN MONIUSZKO**

Sztuczna inteligencja w medycynie – stan aktualny i wyzwania wdrożeniowe

1. Sztuczna inteligencja i uczenie maszynowe

Pojęcie sztucznej inteligencji jest nieostre i zmienia się w czasie. Andrew Moore z Uniwersytetu Carnegie Mellon sztuczną inteligencją nazywa obszar nauki i techniki starający się opracować systemy komputerowe, które zachowują się w sposób, który do tej pory, przypomina ludzką inteligencję. W tym zdaniu ważne jest stwierdzenie „do tej pory”. Pół wieku temu, komputery grające w szachy były uznawane za sztuczną inteligencję. Gdy w roku 1997 komputer Deep Blue pokonał mistrza szachowego Garrego Kasparova zostało to odebrane jako szczytowe osiągnięcie sztucznej inteligencji. Zastosowano wówczas algorytm, który przeszukiwał drzewo możliwych ruchów i wybierał taką kombinację decyzji, która zbliżała gracza do wygranej. Dwadzieścia lat później takie podejście nazwalibyśmy „brutalnym” i niemającym nic wspólnego ze sztuczną inteligencją. Przykład ten pokazuje, że sztuczna inteligencja to raczej ewoluujący cel, do którego dążymy. Do osiągnięcia tego celu wykorzystujemy różne algorytmy, w tym algorytmy uczenia maszynowego.

Uczenie maszynowe jest poddziedziną sztucznej inteligencji. Tom Mitchell z Uniwersytetu Carnegie Mellon w USA w swojej książce pod tytułem *Machine Learning* [1] twierdzi, że obszar naukowy najlepiej definiować przez główne pytanie, na jakie naukowcy z tego obszaru starają się znaleźć odpowiedź. W przypadku uczenia maszynowego pytanie to brzmi: „Jak można zbudować system informatyczny, którego umiejętności będą automatycznie rosły wraz ze wzrostem doświadczenia i jakie podstawowe prawa opisują procesy uczenia?” Przez doświadczenie rozumie się dane dostarczane do algorytmów komputerowych. Uczenie maszynowe to dziedzina, w której pracuje się nad rozwojem programów, które automatycznie zwiększają swoje umiejętności w rozwiązywaniu danego problemu przez doświadczenie zawarte w danych. Uczenie maszynowe może być więc traktowane jako środek do osiągnięcia celu, jakim jest sztuczna inteligencja.

2. Wybrane przykłady zastosowań uczenia maszynowego w medycynie

W niniejszym rozdziale przedstawiono najnowsze osiągnięcia z zakresu zastosowania sztucznej inteligencji w medycynie, chociaż w istocie omawiane przykłady dotyczą

* Dr inż. Konrad Wojdan (konrad.wojdan@pw.edu.pl), Politechnika Warszawska, Instytut Techniki Ciepłej; Transition Technologies Science sp. z o.o.

** Prof. dr hab. Marcin Moniuszko, Uniwersytet Medyczny w Białymstoku

wykorzystania algorytmów uczenia maszynowego. Przedstawione przykłady koncentrują się na odkryciach publikowanych w ciągu ostatnich trzech lat.

2.1. Diagnozowanie raka piersi

Według American Cancer Society (ACS) klinicyści w około 20% przypadków nie wykrywają możliwego do zdiagnozowania raka piersi. Z drugiej strony ACS wskazuje na to, że około 50% kobiet wykonujących co roku mammografię uzyska w ciągu 10 lat fałszywie pozytywny wynik. To skłoniło specjalistów z Google Health i powiązanego DeepMind do stworzenia algorytmu sztucznej inteligencji wykrywającego raka piersi na podstawie mammografii. Razem z Northwestern University, Cancer Research U.K. Imperial Centre i Royal Surrey County Hospital naukowcy przetrenowali Głęboką Sztuczną Sieć Neuronową na zbiorze 25 856 mammografii z Wielkiej Brytanii i 3097 mammografii z USA.

Algorytm okazał się lepszy od wszystkich sześciu radiologów uczestniczących w badaniu. Średnio okazał się o 12,1% skuteczniejszy, biorąc pod uwagę wskaźnik, jakim jest pole pod powierzchnią krzywej ROC. Liczba ta przekłada się na zmniejszenie liczby fałszywie pozytywnych diagnoz o 5,7% w przypadku USA i 1,2% w przypadku Wielkiej Brytanii, oraz zmniejszenie liczby fałszywie negatywnych diagnoz o 9,4% w przypadku USA i 2,7% w przypadku Wielkiej Brytanii.

Badanie to zostało opisane w artykule [2] w „Nature” w styczniu 2020 roku. Stanowi ono dobrą podstawę do dalszych prospektywnych badań klinicznych. Autorzy sami przyznają, że zróżnicowanie populacji pacjentów, jak również zróżnicowanie mammografii, było za małe, aby móc wprowadzić ten algorytm na stałe do praktyki klinicznej. Jest to jednak istotny krok na drodze do wyposażenia radiologów w nowe narzędzia sztucznej inteligencji.

2.2. Diagnozowanie raka płuc

W 2019 roku naukowcy z Google i Northwest University stworzyli algorytm sztucznej inteligencji, który był w stanie wykrywać złośliwe guzki płucne (ang. *malignant lung nodules*) w zdjęciach pochodzących z niskodawkowej tomografii komputerowej klatki piersiowej (ang. *low-dose chest computed tomography* – LDCT) [3]. Wyniki działania algorytmu zostały porównane z diagnozami sześciu radiologów. W przypadku gdy nie dysponowano poprzednimi wynikami tomografii komputerowej, algorytm uzyskiwał wyższą skuteczność niż radiolodzy. W przypadku gdy poprzednie wyniki TK były dostępne, średnia skuteczność algorytmu była podobna do średniej skuteczności radiologów.

Jako jedną z przyczyn sukcesu algorytmu podaje się fakt, że radiolodzy patrzą na płuco jak na sekwencje obrazów 2D. Opisany w pracy algorytm analizuje „w całości” obraz 3D płuca, dzięki czemu ma potencjał, żeby lepiej wychwycić wczesne objawy raka płuc. W sytuacji gdy dostępne są poprzednie wyniki badań TK, algorytm analizuje

jednocześnie zarówno bieżące, jak i poprzednie zdjęcie, co dodatkowo zwiększa skuteczność diagnostyki. Oczywiście autorzy podkreślają, że jest to wstępny etap badań, który wymaga dalszych prac, w szczególności treningu algorytmu na bardziej zróżnicowanej populacji.

2.3. Diagnozowanie na podstawie RTG płuc

Specjaliści z Stanford University opublikowali w listopadzie 2018 roku w PLOS Medicine artykuł na temat algorytmu CheXNet [4]. Algorytm ten potrafi wykryć 14 chorób na podstawie zdjęcia rentgenowskiego płuc. Jakość algorytmu była porównywana z diagnozami dziewięciu radiologów. W przypadku 11 z 14 chorób algorytm miał skuteczność lepszą niż ludzie. Dodatkowo analiza zdjęć RTG zajęła radiologom średnio 3 godziny, podczas gdy algorytm wykonał to zadanie w niecałe 2 minuty.

W przypadku tego badania Głęboka Sztuczna Sieć Neuronowa została wytrenowana na podstawie 112 000 zdjęć rentgenowskich. W przeciwieństwie do poprzednich opisywanych badań, informacja o wystąpieniu choroby nie pochodziła z przeprowadzonych biopsji. W tym przypadku każde zdjęcie opisywało trzech radiologów, w kontekście 14 analizowanych chorób. Radiolodzy wspólnie musieli ustalić, jaka jest ostateczna diagnoza. W tym miejscu należy się czytelnikowi komentarz. Doświadczenia autora niniejszej publikacji wskazują na fakt, że w przypadku braku ostatecznego źródła potwierdzenia diagnozy (np. badania histopatologicznego) bardzo ważne jest, aby podczas opisywania zbioru treningowego wypracować metodologię chroniącą przed ludzkimi błędami.

Pracując nad algorytmem CheXNet, naukowcy planują zebrać i opisać zdjęcia RTG klatki piersiowej z większej liczby ośrodków, tak aby zadbać o większą wariancję w danych i aby przygotować algorytm do prób klinicznych. Może on stanowić wsparcie dla radiologów, chociażby przy wstępnej ocenie i wstępnej kategoryzacji zdjęć RTG.

2.4. Analiza obrazów mózgu

W czerwcu 2018 roku w Pekinie odbyły się pierwsze zawody między lekarzami i sztuczną inteligencją w analizie danych obrazowych mózgu. W zespole lekarzy uczestniczyło 15 uznanych lekarzy, specjalistów w zakresie neuroobrazowania. Wyniki oceniane były przez jury, a wszystkiemu przyglądała się zebrana na sali publiczność. Stającym w szranki z lekarzami algorytmem był BioMind, wytrenowany do diagnozowania guzów mózgu i przewidywania ekspansji krwiaka. Dane do nauki algorytmu zebrano w Beijing Tiantan Hospital, a prawdziwość opisów została potwierdzona badaniami neuropatologicznymi.

W pierwszej rundzie każdy z 15 doświadczonych lekarzy musiał ocenić 15 wyników badań rezonansu magnetycznego celem postawienia lub wykluczenia diagnozy guza mózgu. Każdy z lekarzy miał na to 30 minut (2 minuty na pojedynczy przypadek). W sumie ocenionych zostało 225 wyników badań MR. Lekarze właściwie zdiagnozowali 66%

obrazów, podczas gdy algorytm AI miał o 21% większą skuteczność, stawiając 87% poprawnych diagnoz. Ponadto BioMind potrzebował zaledwie 15 minut, aby przeanalizować wszystkie 225 wyników badań. Algorytm miał przewagę szczególnie przy rzadkich typach nowotworów. W drugiej rundzie inna grupa lekarzy przewidywała ekspansje krwiaka śródmózgowego. Wynik był jednak podobny. Skuteczność lekarzy wyniosła 63%, podczas gdy BioMind właściwie odpowiedział w 83% przypadków.

Wiele podobnych prac zostało opublikowanych w 2020 roku w odniesieniu do raka szyjki macicy, raka skóry i innych typów nowotworów. Jednak sztuczna inteligencja to nie tylko wsparcie diagnostyki w oparciu o analizy obrazów medycznych. Algorytmy uczenia maszynowego analizują sygnały EKG, EEG, strumienie audio ze stetoskopu, jak również klasyczne dane medyczne w połączeniu z danymi „omicznymi”.

2.5. Przewidywanie śmierci na podstawie EKG i problem czarnej skrzynki

W listopadzie 2019 roku na zjeździe American Heart Association ogłoszono, że sztuczna inteligencja jest w stanie rozpoznać ryzyko zgonu na podstawie zapisu elektrycznej aktywności serca z elektrokardiogramu. Brandon Fornwalt z amerykańskiej firmy Geisinger wprowadził do Sztucznej Sieci Neuronowej dane z 1,7 mln zapisów EKG niemal 400 tysięcy osób. Zadaniem sieci było przewidzenie, czy osoba, której EKG jest analizowane, przeżyje kolejny rok. Algorytm miał dodatkowe informacje o wieku i płci pacjenta.

Lekarze na podstawie tych informacji oraz zapisu EKG byli w stanie przewidzieć z prawdopodobieństwem 0.65–0.8 czy pacjent przeżyje kolejny rok. Komputer osiągał wynik powyżej 0.85. Algorytm był w stanie wychwycić pewne niuanse w zapisie EKG, na podstawie których kwalifikował osoby do grupy ryzyka. Były to niuanse na tyle subtelne, że nawet trzech analizujących takie przypadki kardiologów nie było w stanie stwierdzić, jaki wzorec w EKG mógł być przesłanką dla takiej decyzji. Jest to zarówno zaleta, jak i wada algorytmów AI. Z jednej strony algorytmy AI mają potencjał, aby nauczyć ludzi lepszego analizowania zapisów EKG. Z drugiej strony algorytmom bardzo ciężko jest podać podstawy swoich decyzji (problem „czarnej skrzynki”). To powoduje, że lekarze nie mają zaufania do tych algorytmów.

Obecnie trwają intensywne prace nad metodami wyjaśniającymi działanie algorytmów uczenia maszynowego (tzw. XAI – *eXplainable Artificial Intelligence*). Trudno jest wymagać od lekarzy aby podejmowali decyzje w sprawie leczenia konkretnego pacjenta na podstawie wskazań, których nikt nie potrafi wyjaśnić. W pierwszej kolejności trzeba zapewnić transparentność algorytmów sztucznej inteligencji, tak aby zbudować nić zaufania na linii lekarz – maszyna.

2.6. Sztuczna inteligencja wykrywająca zakażenia Covid-19

W październiku 2020 roku naukowcy z MIT udostępnili w Internecie algorytm sztucznej inteligencji wykrywający bezobjawowe zakażenia Covid-19 na podstawie na-

grań kaszlu zarejestrowanych telefonem komórkowym. Badacze zauważyli, że kaszel osób bezobjawowo zakażonych różni się od kaszlu osób zdrowych. Różnice te nie są jednak możliwe do wychwycenia za pomocą ludzkiego ucha. W pracy opublikowanej w [5] naukowcy opisali model przetrenowany na dziesiątkach tysięcy nagrań kaszlu i mowy zebranych przez ochotników za pomocą portalu internetowego. Autorzy użyli wstępnie przetrenowanej Sztucznej Sieci Neuronowej, która wykrywała wczesne objawy choroby Alzheimera. Następnie na podstawie 4000 próbek kaszlu, w tym około połowy zawierającej kaszel pacjentów z potwierdzonym Covid-19, dotrenowali sieć tak, aby wykrywała osoby chore na Covid-19, w tym zakażone bezobjawowo. Na próbie testowej zawierającej 1000 nagrań kaszlu autorzy uzyskali skuteczność na poziomie 98,5% poprawnie zidentyfikowanych osób z Covid-19. W tej grupie skuteczność wykrywania pacjentów bezobjawowych wyniosła 100%. Obecnie autorzy rozwiązania pracują nad udostępnieniem darmowej aplikacji i certyfikacją FDA (U.S. Food and Drug Administration).

2.7. Sztuczna inteligencja na tropie nowych antybiotyków

Wykorzystanie sztucznej inteligencji w medycynie to nie tylko wsparcie diagnostyki badań przesiewowych. Przykładem tego jest projekt, w którym naukowcy z Institute for Medical Engineering and Science MIT (IMES) wykorzystali sztuczną inteligencję do znalezienia potężnego antybiotyku [6]. W 2020 roku naukowcom udało się zidentyfikować antybiotyk, który w testach laboratoryjnych zabija najbardziej problematyczne bakterie, w tym szczepy odporne na wszystkie znane antybiotyki.

Obecne metody poszukiwania nowych antybiotyków są kosztowne, czasochłonne i najczęściej ograniczają się do wąskiego spektrum różnorodności chemicznej. Dlatego też naukowcy z MIT opracowali algorytm uczenia maszynowego, który na podstawie 2335 molekuł o potwierdzonym działaniu antybakteryjnym nauczył się korelować strukturę związków chemicznych z ich cechami, takimi jak zdolność do niszczenia bakterii. Wśród tych cząsteczek było 300 antybiotyków i 800 naturalnych produktów o znanej aktywności biologicznej. Następnie przetrenowany model został wykorzystany do przetrenowania biblioteki zawierającej 6000 związków chemicznych, które są badane w kontekście ludzkich chorób (*Drug Repurposing Hub*). Sto najlepszych wyników zwróconych przez Sztuczną Sieć Neuronową zostało poddanych dalszym testom laboratoryjnym. Jedną z tych molekuł była halicyna. Okazało się, że związek ten miał niską toksyczność dla komórek ludzkich i silną aktywność antybakteryjną. Co więcej, struktura chemiczna halicyny różni się od istniejących antybiotyków. W ten sposób molekula, która była rozpatrywana jako potencjalny lek na cukrzycę, ma szansę zostać nowym, potężnym antybiotykiem.

Na koniec Sztuczna Sieć Neuronowa przeanalizowała bazę zawierającą 107 milionów molekuł. Algorytm potrzebował na to 3 dni. Z 23 zidentyfikowanych przez sztuczną

inteligencję cząsteczek naukowcy potwierdzili antybakteryjne działanie w przypadku 8 molekuł. Dwie z tych cząsteczek to związki silnie bakteriobójcze.

2.8. Przewidywanie struktury przestrzennej białek

Dzięki algorytmom sztucznej inteligencji w 2020 roku dokonał się także przełom w innym istotnym z punktu widzenia rozwoju medycyny obszarze – przewidywaniu struktury przestrzennej białek. Wiele istotnych problemów badawczych w medycynie (w tym np. projektowanie nowych leków) bazuje na prawidłowym przewidzeniu wyniku związania się białek. Ostateczny kształt białka wpływa na funkcję, jaką to białko pełni w organizmie. W 1972 roku zdobywca Nagrody Nobla Christian Anfinsen powiedział, że w teorii sekwencja aminokwasów białka powinna w pełni determinować jego kształt. Niestety, ilość możliwości, w jakich jednowymiarowa sekwencja aminokwasów może być zmieniona w trójwymiarowy kształt białka, jest ogromna.

Do 2016 roku najlepsze algorytmy były w stanie przewidzieć strukturę 3D białka z GDT wynoszącym ok. 40. Miara GDT (*Global Distance Test*) może przyjmować wartości od 0 do 100 i mówi o procentowej liczbie aminokwasów, dla których wartość różnicy pomiędzy przewidywaną i rzeczywistą strukturą wynosi mniej niż pewien przyjęty próg dokładności. AlphaFold 2, najnowszy algorytm sztucznej inteligencji stworzony w 2020 roku, uzyskuje wyniki na poziomie 92,4 GDT. Średni błąd przewidywania kształtu białka wynosi 1,6 Angstrema i jest to porównywalne ze średnicą atomu. Tak dokładne modelowanie kształtu białek to przełom w dziedzinie biologii molekularnej. Na uwagę zasługuje fakt, że dzięki algorytmowi AlphaFold 2 będzie można lepiej poznać białka błonowe, których krystalizacja jest trudna. Praca na temat algorytmu została opublikowana w *Nature* w 2020 roku [7].

2.9. EPISTOP – prewencyjne leczenie padaczki

EPISTOP to projekt koordynowany przez Instytut „Pomnik – Centrum Zdrowia Dziecka”, realizowany razem z 16 partnerami z Europy, USA i Australii. Celem projektu było zbadanie przyczyn i przebiegu padaczki u dzieci chorych na stwardnienie guzowate [8]. Noworodki ze stwardnieniem guzowatym mają wysokie ryzyko rozwoju epilepsji, niepełnosprawności intelektualnej i autyzmu. Leczenie prewencyjne istnieje, ale dotychczas obawiano się skutków ubocznych.

W projekcie zebrano 33 TB danych od 101 małych pacjentów. Były to dane genomiczne, transkryptomiczne, proteomiczne, metabolomiczne, kliniczne, obrazy MRI, badania EEG i wywiady psychologiczne. W oparciu o klasyczne metody statystyczne wykazano skuteczność leczenia prewencyjnego. Dzięki wykorzystaniu uczenia maszynowego stworzono też algorytm identyfikujący potencjalne biomarkery rozwoju padaczki. Tak ogromna ilość danych zarówno klinicznych, jak i omicznych w naturalny sposób otwiera drogę

do wykorzystania algorytmów sztucznej inteligencji, szczególnie w analizach eksploracyjnych ukierunkowanych na poszukiwania biomarkerów. W medycynie szczególnie istotny bywa problem małej liczby pacjentów uczestniczących w badaniu w stosunku do dużej ilości klinicznych i molekularnych zmiennych opisujących pojedynczego pacjenta. Wówczas bardzo ważne jest umiejętne połączenie algorytmów redukujących przestrzeń cech z algorytmami uczenia nadzorowanego oraz dbanie o to, aby modele posiadały wysoki stopień generalizacji (unikanie problemu „przetrenowania”).

2.10. Sztuczna inteligencja wspierająca telediagnostykę

W związku z rozwojem telemedycyny na rynku pojawiają się rozwiązania wspomagające proces zdalnej diagnostyki, w szczególności przeprowadzanej w warunkach domowych. Na uwagę zasługują rozwiązania opracowywane przez polskie firmy HigoSense, StethoMe czy MedApp. Algorytmy sztucznej inteligencji mogą pomóc przeanalizować zebrany materiał diagnostyczny, aby wesprzeć lekarza w postawieniu diagnozy. Należy jednak zwrócić uwagę na jeszcze jedno ważne zagadnienie, w którym także potrzebne jest wsparcie maszyny. Otóż sam proces wykonania badania i zebrania odpowiedniego materiału diagnostycznego może nie być rzeczą prostą dla osoby niewykwalifikowanej. Dlatego też firma HigoSense wyposażyła swoje urządzenie telemedyczne w algorytmy sztucznej inteligencji, opracowane we współpracy z Transition Technologies Science, które pomagają użytkownikowi (najczęściej rodzicowi) przeprowadzić badanie. Zadaniem algorytmu SI jest analiza strumienia wideo z obrazami gardła i ucha (cyfrowy otoskop) oraz analiza osłuchów serca i płuc (cyfrowy stetoskop) celem stwierdzenia, czy zarejestrowany obraz lub dźwięk jest „diagnostyczny”. Rodzic dostaje wskazówki, czy właściwie wykonuje badanie i czy zebrał już odpowiedni materiał, który jest dla lekarza wystarczający w kontekście postawienia diagnozy.

3. Sztuczna inteligencja w medycynie

– gdzie jesteśmy i jakie stoją przed nami wyzwania?

W październiku 2019 roku w „Lancet Digital Health” opublikowany został artykuł *Human versus machine in medicine: can scientific literature answer the question?* [9]. Przeanalizowano w nim 25, dobrze udokumentowanych prac naukowych porównujących skuteczność diagnozy postawionej przez człowieka i maszynę. Prace te dotyczyły 13 różnych chorób, z czego najliczniej reprezentowane były rak piersi i rak skóry. Wnioski z przeprowadzonej analizy były takie, że faktycznie występujące schorzenia komputery diagnozowały poprawnie w 87% przypadkach, zaś ludzie w 86% przypadkach. Przypadki zdrowe były poprawnie diagnozowane przez algorytmy ze skutecznością 91%, zaś przez lekarzy ze skutecznością 93%. Tak więc wyniki sztucznej inteligencji i ludzi są do siebie zbliżone.

W artykule podkreśla się, że analizowane prace często oderwane były od praktyki klinicznej. Tylko w przypadku czterech publikacji lekarze mieli wgląd w pełne dane o pacjencie, tak jak to ma miejsce na co dzień w szpitalu. W większości prac lekarze oceniali po prostu obrazy medyczne w oderwaniu od szerszego kontekstu wiedzy o pacjencie. Dodatkowo podkreślany jest fakt, że jakość prac publikowanych w dziedzinie sztucznej inteligencji jest niska. Spośród tysięcy prac poświęconych tej tematyce, tylko 25 miało na tyle wysoką jakość, aby mogły zostać uwzględnione w zestawieniu.

Ostatecznie autor artykułu formułuje wniosek, że na podstawie istniejących prac można powiedzieć, że dobrze wytrenowana sztuczna inteligencja nie jest gorsza od ludzkich specjalistów. Nie oznacza to, że jest lepsza. Może być ona jednak bardzo pomocna tam, gdzie dostęp do służby zdrowia jest utrudniony lub gdzie występują duże kolejki do specjalistów.

Podsumowując, w obszarze diagnostyki, w szczególności w zakresie diagnostyki obrazowej, algorytmy sztucznej inteligencji dorównują lub przewyższają ludzkich specjalistów, jeśli chodzi o trafność stawianych diagnoz. Trzeba tu jednak wyraźnie podkreślić, że są to najczęściej wyniki badań retrospektywnych, przeprowadzonych na wąskiej populacji, w zakresie której algorytmy były trenowane. Trudno więc mówić o tym, aby algorytmy sztucznej inteligencji mogły w obecnym stanie wejść do codziennej praktyki klinicznej. Stan rozwoju tych metod pozwala jednak na to, aby z optymizmem wejść w fazę trenowania algorytmów na dużej i zróżnicowanej populacji oraz przeprowadzić prospektywne badania kliniczne.

Algorytmy sztucznej inteligencji potrafią bardzo szybko przeanalizować dużą liczbę danych. Mogą więc wspomóc ludzi wszędzie tam, gdzie w procesie diagnostyki tworzą się wąskie gardła. Komputery mogą dokonywać wstępnej oceny i selekcji pacjentów, udostępniając lekarzowi wstępnie ocenione wyniki osób wymagających najszybszej pomocy. Sztuczna inteligencja to nie tylko diagnostyka. Jak zostało pokazane, komputery mogą pomóc w doborze terapii, w projektowaniu nowych leków, a także w rozwoju biologii molekularnej.

Metody SI nie zastąpią inteligencji ludzkiej, tylko będą stanowiły rozszerzenie zdolności człowieka. Komputery wyposażone w SI będą narzędziami w rękach specjalistów. Dzięki nim praca lekarzy będzie skuteczniejsza i bardziej wydajna. Już dzisiaj pojawiają się głosy, że o ile w różnych testach skuteczność diagnoz ludzkich i maszynowych jest porównywalna, to wykorzystanie potencjału obydwu rodzajów inteligencji daje najlepsze wyniki. We wstępnych testach najlepsze diagnozy stawiała nie sama maszyna, lub sam człowiek, a lekarz mogący wykorzystać podpowiedzi algorytmów SI.

Współpraca na linii człowiek – maszyna wymaga jednak zaufania. Obecnie algorytmy uczenia maszynowego to w większości czarne skrzynki. Trwają jednak prace nad tzw. wyjaśnianą sztuczną inteligencją. Lekarze muszą wiedzieć, na jakiej podstawie algorytm

podjął taką, a nie inną decyzję. Dodatkowo, wiedząc, że algorytmy SI z czasem wyjdą z laboratoriów komputerowych i będą towarzyszyły lekarzom w codziennej pracy, trzeba zadbać o to, aby wprowadzić mechanizmy stale weryfikujące jakość tych algorytmów. Należy pamiętać, że maszyny będą stale uczyć się od swoich ludzkich trenerów-lekarzy. Z powodu różnych błędów może się zdarzyć i tak, że dane używane do nauki będą niskiej jakości. Właściwie dobrane mechanizmy weryfikujące jakość algorytmów powinny wówczas wychwycić pogarszającą się skuteczność algorytmów sztucznej inteligencji. Jest to też element zwiększający zaufanie człowieka do maszyny.

Jedną z barier związanych z rozwojem algorytmów sztucznej inteligencji jest dostęp do dobrze oznaczonych i zweryfikowanych danych uczących. Danych tych powinno być stosunkowo dużo i powinny one dotyczyć najlepiej maksymalnie zróżnicowanej populacji. Niestety opisywanie danych na potrzeby nauki algorytmów to często żmudny i długotrwały proces. Wiedząc o tym, warto na bieżąco w sposób ustrukturyzowany gromadzić dane, bo chcąc nie chcąc w niedalekiej przyszłości będziemy z tego typu narzędzi korzystać, więc w interesie lekarzy i pacjentów jest, aby douczać maszyny-pomocników przy użyciu dobrej jakości danych uczących.

Ostatnią przeszkodą techniczną we wdrożeniu algorytmów sztucznej inteligencji jest też problem bieżącego dostępu do danych zgromadzonych w systemach szpitalnych. Aby medycyna mogła się rozwijać, placówka medyczna powinna mieć łatwy, szybki i tani dostęp do danych medycznych. Tak niestety nie jest. Na szczęście na rynku pojawiły się produkty informatyczne „otwierające” systemy szpitalne (jak np. system Med-Stream Designer [10]), a to z kolei pozwala na wdrożenie algorytmów sztucznej inteligencji bazujących na bieżących danych pacjenta.

4. Centrum Sztucznej Inteligencji w Medycynie Uniwersytetu Medycznego w Białymstoku

Poszukiwanie rozwiązań diagnostycznych i terapeutycznych wykorzystujących dane medyczne wysokiej jakości leżało u podstaw stworzenia Centrum Sztucznej Inteligencji w Medycynie na Uniwersytecie Medycznym w Białymstoku (C-SIM UMB). Celem C-SIM UMB było i jest: 1) stworzenie w Polsce silnego centrum wiedzy i rozwoju w zakresie tworzenia tzw. *big data* oraz rozwoju narzędzi sztucznej inteligencji w medycynie, łączącego wysiłki i działania klinicystów, bioinformatyków, specjalistów od biobankowania, ekspertów w genomice i innych technikach badawczych; 2) stworzenie oryginalnych formatów baz w oparciu o unikalne dane medyczne i omiczne; 3) stworzenie systemu udostępniania wysokiej jakości danych biomedycznych dla polskich zespołów analityków sztucznej inteligencji związanych tak z nauką, jak i z przemysłem; 4) zwiększenie skuteczności terapii i lepsze ekonomicznie/bardziej efektywne wydanie funduszy na ochronę zdrowia w Polsce.

Twórcy C-SIM UMB wyszli z założenia, że stworzenie w Polsce rzeczywiście innowacyjnych, unikalnych rozwiązań/algorytmów sztucznej inteligencji nie będzie możliwe bez dostępu do unikalnych zbiorów danych. Nowe narzędzia sztucznej inteligencji w medycynie okażą się tylko tak dobre, jak dokładne i obszerne będą zbiory danych, na których one będą się opierać. A tych unikalnych, naprawdę kompleksowych zbiorów danych w medycynie na ogół nie dostaje się za darmo, a ich wykupienie jest najczęściej bardzo kosztowne. W związku z powyższym UMB postawił na rozwiązania, które doprowadzą do stworzenia wysokiej jakości zbiorów danych na miejscu. Tym bardziej, że w przypadku chorób i wywołujących je zmian genetycznych dane pochodzące (na przykład) od pacjentów onkologicznych z innych części świata nie muszą odzwierciedlać cech nowotworu, na który chorują pacjenci w Polsce. Generalnie, nie ma raczej większych szans na konkurencyjne rozwiązania w dziedzinie sztucznej inteligencji w medycynie bez dostępu do własnych, kompleksowych zbiorów danych dotyczących danej/danych jednostek chorobowych.

UMB jest jedną z nielicznych instytucji w tej części Europy, która tak unikalne jakościowo zbiory danych tworzy, szczególnie w obszarze chorób nowotworowych i innych chorób cywilizacyjnych. Dotychczas było to możliwe dzięki przemyślanemu wykorzystaniu dostępnych funduszy i stworzeniu specjalnych wielodyscyplinarnych zespołów zajmujących się w bezprecedensowy i niespotykany w zbyt wielu miejscach na świecie sposób zbieraniem niesłychanie dokładnych danych dotyczących pacjenta (od analizy wszystkich genów, zobrazowania anatomii i funkcji całego organizmu poprzez szczegółową analizę stylu życia i jego otoczenia). Stworzono unikalne wielodyscyplinarne zespoły badawcze łączące medyków i bioinformatyków pracujących nad tworzeniem nowych rozwiązań sztucznej inteligencji w chorobach cywilizacyjnych. To, co wyróżnia C-SIM UMB, to właśnie funkcjonowanie wszystkich kluczowych elementów tej układanki w jednym i tym samym miejscu. Począwszy od zgodnego z najwyższymi światowymi standardami bankowania próbek materiału biologicznego, przez niezwykle dokładną, ponadstandardową analizę sytuacji klinicznej pacjenta i skończywszy na wykonaniu wielu kompleksowych, unikalnych analiz wielkoskalowych (nie tylko genomicznych). Problem polega na tym, że w związku z olbrzymimi wymogami logistycznymi dramatycznie mała liczba ośrodków naukowych na świecie jest w stanie generować tak złożone bazy danych, które łączyłyby w sobie dane badań genomu ze szczegółowymi danymi o zdrowiu pacjenta i jego otoczeniu. Stworzenie takiego zbioru danych wymaga bowiem niezwyklej determinacji, konsolidacji wielu złożonych aktywności i procedur oraz dyscypliny logistycznej, osiągnięcie których staje się tym trudniejsze, z im większym ośrodkiem mamy do czynienia. W UMB podjęto skoordynowane działania służące zbieraniu wyjątkowo zróżnicowanego i bogatego zbioru danych od poszczególnych pacjentów. Daje to niespotykane zbyt często we współczesnym świecie możliwości stworzenia plat-

formy badawczej będącej podstawą do rozwoju nieskończonej ilości rozwiązań bioinformatycznych opartych na systemach sztucznej inteligencji. Dzięki skokowemu rozwojowi coraz doskonalszych, ale i coraz bardziej dostępnych technologii medycznych świadomość konieczności wykonywania badań genomu staje się coraz bardziej powszechna. Koszty analizy genomu, które jeszcze przed kilkunastu lat sięgały dziesiątków milionów dolarów, obecnie spadły do kilku, kilkunastu tysięcy złotych. Z wymienionych wyżej powodów liczba analiz genomowych na świecie rośnie w błyskawicznym tempie. Dość szybko jednak znawcy rynku nowoczesnych technologii medycznych zorientowali się, że same wyniki badań genomowych, choć niezwykle cenne, nie niosą w sobie istotnej wartości medycznej i rynkowej, jeśli brakuje im szczegółowej korelacji z oceną stanu zdrowia pacjenta, historią jego choroby, odpowiedzią na stosowane leczenie itp. Biorąc to pod uwagę, z całą pewnością można stwierdzić, że kontynuacja podobnych badań w tej dziedzinie, a przede wszystkim ich rozciągnięcie na inne jednostki chorobowe sprawią, że atrakcyjność naukowa i wdrożeniowa takich rozwiązań będzie coraz większa.

Bibliografia

- [1] Mitchell T.M. (1997), *Machine Learning*, McGraw-Hill, New York.
- [2] McKinney S.M., Sieniek M., Godbole V., Godwin J., Antropova N., Ashrafian H., Back T., Chesus M., Corrado G.S., Darzi A., Etemadi M., Garcia-Vicente F., Gilbert F.J., Halling-Brown M., Hassabis D., Jansen S., Karthikesalingam A., Kelly C.J., King D., Ledsam J.R., Melnick D., Mostofi H., Peng L., Reicher J. J., Romera-Paredes B., Sidebottom R., Suleyman M., Tse D., Young K.C., Fauw J.D., Shetty S. (2020). *International evaluation of an AI system for breast cancer screening*. Nature, 577(7788), 89–94. <https://doi.org/10.1038/s41586-019-1799-6>
- [3] Ardila D., Kiraly A.P., Bharadwaj S., Choi B., Reicher J.J., Peng L., Tse D., Etemadi M., Ye W., Corrado G., Naidich D.P., Shetty S. (2019). *End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography*. Nature medicine, 25(6), 954–961. doi: 10.1038/s41591-019-0447-x
- [4] Rajpurkar P., Irvin J., Ball R.L., Zhu K., Yang B., Mehta H., Duan T., Ding D., Bagul A., Langlotz C.P., Patel B.N., Yeom K.W., Shpankaya K., Blankenberg F.G., Seekins J., Amrhein T.J., Mong D.A., Halabi S.S., Zucker E.J., Ng A.Y., Lungren M.P. (2018). *Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists*. PLoS medicine, 15(11), e1002686. <https://doi.org/10.1371/journal.pmed.1002686>
- [5] Laguarda J., Hueto F., Subirana B. (2020), *COVID-19 Artificial Intelligence Diagnosis Using Only Cough Recordings*. IEEE Open Journal of Engineering in Medicine and Biology, t. 1, s. 275–281, doi: 10.1109/OJEMB.2020.3026928.
- [6] Stokes J.M., Yang K., Swanson K., Jin W., Cubillos-Ruiz A., M. Donghia N., MacNair C.R., French S., Carfrae L.A., Bloom-Ackermann Z., Tran V.M., Chiappino-Pepe A., Badran A.H., Andrews I.W., Chory E.J., Church G.M., Brown E.D., Jaakkola T.S., Barzilay R., Collins J.J. (2020). *A Deep Learning Approach to Antibiotic Discovery*. Cell, 180 (4), 688–702. e13, ISSN 0092-8674.

- [7] Senior A.W., Evans R., Jumper J., Kirkpatrick J., Sifre L., Green T., Qin C., Žídek A., Nelson A., Bridgland A., Penedones H., Petersen S., Simonyan K., Crossan S., Kohli P., Jones D.T., Silver D., Kavukcuoglu K., Hassabis D. (2020). *Improved protein structure prediction using potentials from deep learning*. *Nature*, 577(7792), 706–710. <https://doi.org/10.1038/s41586-019-1923-7>
- [8] Kotulska K., Kwiatkowski D.J., Curatolo P., Weschke B., Riney K., Jansen F., Feucht M., Krsek P., Nabbout R., Jansen A.C., Wojdan K., Sijko K., Głowacka-Walas J., Borkowska J., Sadowski K., Domańska-Pakiela D., Moavero R., Hertzberg C., Hulshof H., Scholl T., Benova B., Aronica E., de Ridder J., Lagae L., Jóźwiak S., EPISTOP Investigators (2021). *Prevention of Epilepsy in Infants with Tuberous Sclerosis Complex in the EPISTOP Trial*. *Annals of neurology*, 89(2), 304–314. <https://doi.org/10.1002/ana.25956>
- [9] Cook T.S. (2019). *Human versus machine in medicine: can scientific literature answer the question?* *The Lancet. Digital Health*, 1(6), e246–e247. [https://doi.org/10.1016/S2589-7500\(19\)30124-4](https://doi.org/10.1016/S2589-7500(19)30124-4)
- [10] Raport „Innowacyjna onkologia. Potrzeby. Możliwości. System” red. B. Więckowska, A. Maciejczyk, PZWL Wydawnictwo Lekarskie i Polskie Towarzystwo Onkologii, 2020.

Artificial Intelligence in medicine – current status and implementation challenges

Artificial Intelligence begins to play an increasingly important role in medicine, in particular in diagnostics, therapy selection and drug design. This article shows how the latest machine learning algorithms support the work of physicians and pharmacists. However, the effective implementation of Artificial Intelligence methods in everyday medical practice requires overcoming a number of barriers. These challenges are discussed in the article. The objectives and functioning of the Artificial Intelligence Center in Medicine of the Medical University of Białystok were also discussed, as an example of Polish contribution to the development of the latest computer algorithms supporting diagnostics and therapy.

Key words: Artificial Intelligence, Machine Learning, medicine, diagnostics