



UNIwersytet Medyczny im. Karola Marcinkowskiego w Poznaniu

Katedra Radiologii Ogólnej i Zabiegowej

Ul. DŁUGA 1/2
61-848Poznań

Kierownik: *Prof. dr hab. med. Robert Juszkat*
sekretariat@radiologia.edu.pl

tel. 061 8549-280

fax 061 8549-083

e-mail:

Poznań, 17 lipca 2020

Recenzja rozprawy doktorskiej

M.D. Charlie Hamm

pt. „Development of an Interpretable Liver Tumor Diagnosis Tool
using Deep Learning”

Przedstawiona mi do oceny rozprawa doktorska oparta jest na cyklu publikacji naukowych dotyczących diagnozowania nowotworów wątroby przy użyciu badań radiologicznych rezonansu magnetycznego z wykorzystaniem uczenia głębokiego (ang. deep learning), które jest podkategorią uczenia maszynowego i ma na celu współpracę wybranej sieci neuronowej z lekarzami radiologami lub dążenie do samodzielnego diagnozowania przypadków klinicznych na podstawie badań radiologicznych. Doktorant podjął się zgłębienia ciekawego oraz aktualnego tematu. Temat pracy pokazuje niezwykle szybki rozwój technologiczny wykorzystywany m.in. w branży medycznej i nasuwa niezwykle istotne pytanie: Czy kiedykolwiek maszyna będzie w stanie współpracować z człowiekiem lub być na tyle inteligentna, aby zastąpić lekarza radiologa w procesie opisywania obrazów radiologicznych? Na podstawie przeanalizowanej rozprawy oraz załączonych publikacji naukowych doktoranta, widać głębokie wieloletnie zainteresowanie przedstawionym tematem oraz jego dobrą znajomość.

Praca składa się z krótkiego wstępu, w którym doktorant wyjaśnia zagadnienie jednostki chorobowej jaką jest rak wątrobowokomórkowy oraz przedstawia potencjalne znaczenie połączenia badań obrazowych z wykorzystaniem deep learningu, w procesie diagnostycznym choroby. Doktorant przedstawił wyniki analiz własnych z wykorzystaniem systemu standaryzacji LI-RADS (ang. Liver Imaging Reporting and Data System), który ma za zadanie gromadzenie danych, opisywanie oraz interpretację otrzymanych wyników, uzyskanych podczas badań obrazowych tomografii komputerowej oraz rezonansu magnetycznego u pacjentów, którzy znajdują się w grupie ryzyka zachorowania na nowotwór wątrobowokomórkowy. W badaniu własnym, doktorant skupił się na analizie wyników badań rezonansu magnetycznego. Doktorantowi zależało na rozwoju systemu maszynowego o wysokiej dokładności, porównywalnej z diagnozą lekarską.

Doktorant przedstawił jeden ogólny cel pracy. Brakuje pełnego wyszczególnienia elementów, na których skupił się doktorant.

Kolejnym elementem jest opis materiału i metod rozprawy. Doktorant przedstawił wyniki uzyskane na podstawie jednoośrodkowego retrospektywnego badania, w którym rozwinięto system klasyfikacji nowotworu wątroby z wykorzystaniem struktury sieci CNN (ang. convolutional neural network; pl. konwulcyjne sieci neuronowe) oraz wdrożył własny algorytm, dedykowany analizie ukrytych powłok CNN w modelu antagonistycznym. W opisie grupy badanej brakuje informacji na temat ilości osób, których badania obrazowe zostały wykorzystane w badaniu własnym. Informacja podana jest dopiero na kolejnych stronach w tabeli i wynosi 296. Przedmiotem badania były obrazy MRI jamy brzusznej wykonane w latach 2010-2017, pobrane z systemu PACS. Nie podano konkretnej nazwy miejsca (szpitala), z którego czerpano dane. Obrazy zostały wytypowane na podstawie następujących słów kluczowych: naczyniak krwionośny jamisty, ogniskowy rozrost guzkowy, torbiel prosta, wewnątrzwątrobowy rak dróg żółciowych, przerzutowy rak jelita grubego oraz rak wątrobowokomórkowy.

Następnie, doktorant przedstawił protokół analizowanych badań MRI oraz opisał zastosowany model uczenia maszynowego CNN. Dokładność oceny badań diagnostycznych została porównana pomiędzy modelem CNN a 2 specjalistami z dziedziny radiologii z 39 oraz 7-letnim doświadczeniem. Porównano 60 w pełni

zaślepionych badań (po 10 z każdej grupy chorobowej). Wykonanie zadania określono w 2 kategoriach: zdolność klasyfikacji nowotworu w 6 podanych kategoriach oraz 3 bardziej obszernych kategoriach, podanych przez LI-RADS: LR-1 reprezentującą nowotwory łagodne, LR-5 wskazującą wyłącznie raka wątrobowokomórkowego oraz LR-M rozpoznającą nowotwory inne niż HCC.

Opisy szczegółowej interpretacji wyników otrzymanych na podstawie systemu maszynowego zostały przedstawione w publikacjach naukowych doktoranta. Do pracy wykorzystano narzędzia statystyczne pozwalające określić dokładność, czułość i swoistość wykorzystanych metod. Wyniki badań również zostały przedstawione wcześniej w publikacjach naukowych.

Model głębokiego uczenia wykazał średnią precyzyjność testów na poziomie 98% specyficzności, 92% czułości oraz 92% dokładności. W badaniu mającym na celu porównanie skuteczności diagnostycznej pomiędzy wybranymi technologiami (Reader Study) względem 6 wybranych kategorii, wykazano skuteczność modelu na poziomie 90% (rozpoznano 55/60 zmian), podczas gdy skuteczność dwóch radiologów osiągnęła kolejno poziom 80% oraz 85%. Dla 3 kategorii wybranych przez program, skuteczność wynosiła 92% dla modelu oraz 88% dla radiologów.

Skuteczność CNN oceniona została wysoka dla nieskomplikowanych przypadków, jak np. hiperwzmocnienie w fazie tętnicznej czy wzmocniona masa w fazie opóźnionej. Skuteczność była mniejsza w bardziej kompleksowych przypadkach, np. grudkowatość etc. W związku z tym, skuteczność modelu można uznać za zawyżoną, ponieważ wysoka skuteczność na przykładach prostych, przekłada się na prawie całkowity brak skuteczności w trudniejszych przypadkach (np. 7% skuteczności w rozpoznawaniu heterogenicznych zmian ICC).

Kolejnym elementem pracy jest obszerna dyskusja. Doktorant podkreśla wysoką skuteczność i dokładność ocen zmian nowotworowych w wątrobie za pomocą uczenia maszynowego na podstawie obrazów MR. Co więcej, lek. Hamm dodaje, iż w wielu przypadkach, system osiągnął lepsze wyniki niż oceniający badania lekarze radiolodzy.

Jednym z atutów pracy jest fakt, że dotychczasowe badania tego modelu przeprowadzono na podstawie analiz obrazów badań ultrasonograficznych oraz tomografii komputerowej. Praca doktorska skupia się wyłącznie na ocenie obrazów rezonansu

magnetycznego, co pozwala na jej dość wnikliwą ewaluację. Autor podkreśla ograniczenia modelu, takie jak konieczność ustalenia konkretnych grup klasyfikacyjnych, bez których system nie będzie w stanie sam sklasyfikować zmiany. Model wykazał dość wysoką średnią skuteczność w przypadku 6 grup klasyfikacyjnych (92%-96%). Co jednak w przypadku znacznie większej ilości grup klasyfikacyjnych? Czy istnieje ryzyko, że skuteczność zmaleje w przypadku pojawienia się dużo szerszej skali oceny oraz konieczności zastosowania dużo bardziej rozbudowanego i dokładnego algorytmu? Doktorant stwierdza, iż system ma większe szanse na zdiagnozowanie HCC niż lekarz radiolog w przypadku, gdy obraz jest niejednoznaczny. Dlatego też, model maszynowy może służyć jako pomoc w przypadkach, gdy lekarz radiolog wskazuje wątpliwości dotyczące interpretacji uzyskanych obrazów. Na podstawie uwag przekazanych przez system, lekarz radiolog może przeanalizować poszczególne elementy badania.

W kwestii ograniczeń badania własnego doktorant stwierdza, że należy przeprowadzić wielośrodkowe badania z wykorzystaniem większej ilości różnorodnych obrazów MRI, aby dokładnie określić skuteczność stosowania rozwiązania tego typu modelu maszynowego. Jak podkreśla doktorant, w pracy użyto wyłącznie obrazy zawierające charakterystyczne cechy HCC, wykluczając bardziej urozmaicone oraz kompleksowe przypadki, co sprawia, że LI-RADS może zostać zastosowane wyłącznie u osób z wysokim ryzykiem HCC.

Moim zdaniem, biorąc pod uwagę otrzymane wyniki, można zdecydowanie wyciągnąć wnioski, iż uczenie maszynowe może być pomocne w niektórych nieskomplikowanych działaniach diagnostycznych, lecz nie jest w stanie ocenić kompleksowych przypadków oraz nie jest niezawodne, w pełni skuteczne i nie jest w stanie zastąpić działań lekarza radiologa.

Jednym z mankamentów pracy jest brak oznaczenia numerów stron oraz przypisania stron w spisie treści, co odrobinę utrudnia znalezienie poszczególnych elementów w pracy. W pracy pojawiają się nieliczne błędy edytorskie oraz językowe. W pracy użyto 50 anglojęzycznych publikacji, w celu opracowania części teoretycznej pracy.

Podsumowując, praca napisana jest przemyślanie i starannie. Na uwagę zasługuje fakt, iż badania przytoczone przez doktoranta zostały już dostrzeżone i docenione poprzez publikację ich w znanym i cenionym czasopiśmie naukowym. Na podstawie przedstawionych informacji można stwierdzić, iż autor wykazał się szeroką znajomością tematu, a układ pracy jest poprawny, logiczny oraz zgodny z tematem. Autor poprawnie wykorzystał wybrane metody badawcze i wyciągnął wnioski pozwalające odpowiedzieć na pytania postawione przez autora, w celu oceny wybranego zagadnienia.

Według mojej opinii, przedstawiona rozprawa doktorska spełnia wszystkie formalne wymagania stawiane dysertacjom na stopień doktora nauk medycznych, które znaleźć można w Ustawie o stopniach naukowych i tytułach naukowych.

W związku z powyższym, przedkładałam wniosek do Rady Wydziału Lekarskiego Pomorskiego Uniwersytetu Medycznego w Szczecinie o dopuszczenie rozprawy M.D. Charliego Hamma do dalszych etapów przewodu doktorskiego.

Prof. dr hab. Robert Juszkat
KIEROWNIK
KATEDRY RADIOLOGII OGÓLNEJ I ZABIEGOWEJ
Prof. dr hab. Robert Juszkat

