**Charlie Hamm POLISH ABSTRACT**

**Cele**

Celami tego badania było: (i) opracowanie wysokiej dokładności systemu głębokiego uczenia się do oceny i klasyfikacji zmian w wątrobie przy pomocy rezonansu magnetycznego z kontrastem,
z możliwością oceny podjętej decyzji przez lekarza oraz,(ii) walidacja tego systemu przez porównanie jego wyników diagnostycznych z wynikami uzyskanymi przez lekarzy radiologów.

**Metody**

Badanie objęło 296 pacjentów z 494 zmianami chorobowymi wątroby podzielonymi na sześć kategorii. Zmiany zostały zidentyfikowane za pomocą wielofazowego MRI i podzielone na zestawy treningowe (n=434) i testowe (n=60). Dostępne techniki multiplikacji obrazu zostały wykorzystane w celu zwiększenia liczby próbek szkoleniowych do 43 400. Ten zestaw szkoleniowy wprowadzono do stworzonej na zamówienie sieci neuronów konwulsyjnych (CNN), składającej się z trzech warstw konwulsyjnych z powiązanymi prostymi jednostkami liniowymi, dwóch maksymalnych warstw zbiorczych oraz dwóch w pełni połączonych warstw.

Optymalizator Adama został użyty w celu szkolenia. Maksymalnie cztery kluczowe cechy na każdą zmianę chorobową zostały przypisane do podzbioru każdej klasy zmiany, dodatkowo zastosowano algorytm post-hoc do wnioskowania o obecności tych cech w zestawie testowym na podstawie wzorów aktywacji (wytrenowanego) modelu CNN. Walidacja CNN została przeprowadzona poprzez porównanie wyników diagnostycznych CNN z wynikami dwóch specjalistów radiologii. Zostało to przeprowadzone w ramach walidacji krzyżowej Monte Carlo, a wyniki CNN na identycznym, niewidocznym zestawie testowym zostały porównane z wynikami radiologów. Wygenerowane zostały mapy cech wyróżniające regiony na oryginalnym obrazie, które odpowiadały poszczególnym cechom. Następnie do każdej zidentyfikowanej cechy przypisano ocenę istotności, oznaczającą względne znaczenie danejcechy dla przewidywanej klasyfikacji zmiany.

**Wyniki**

Interpretowalny system głębokiego uczenia się (DL) wykazał 92% czułości (Sn), 98% specyficzności (Sp) i 92% dokładności. Wydajność zestawu testowego w pojedynczym badaniu wykazała średnio 90% Sn i 98% Sp w sześciu typach zmian, w porównaniu do średnio 82,5% Sn i 96,5% Sp dla radiologów. Radiolodzy uzyskali Sn na poziomie 60%-70% do klasyfikacji raka wątrobowokomórkowego, natomiast system DL uzyskał Sn na poziomie 90%. Dla szczególnego przypadku klasyfikacji HCC CNN uzyskał obszar charakterystyki pracy pod krzywą 0,992. Czas obliczeniowy na jedną zmianę wynosił 5,6 milisekundy.

Dodatnia wartość predykcyjna i Sn w identyfikacji prawidłowych cech radiologicznych występujących w każdej badanej zmianie wynosiły odpowiednio 76,5% i 82,9%, podczas gdy 12% zmian było źle sklasyfikowanych; te źle sklasyfikowane zmiany częściej prowadziły do błędnej identyfikacji cech niż prawidłowo sklasyfikowane (60,4% vs 85,6%). Oryginalne woksle obrazowe przyczyniające się do każdej funkcji obrazowania były spójne z wygenerowanymi mapami cech, a w każdej klasie najbardziej znaczące kryteria obrazowania były odzwierciedlone przez ich odpowiednie oceny istotności cech.

**Wniosek**

W pracy przedstawiono rozwój "interpretowalnego" prototypu systemu głębokiego uczenia się, którego dokładność przewyższa dokładność radiologów w klasyfikowaniu zmian w wątrobie na MRI wzmocnionym kontrastem, przy jednoczesnym oświetleniu procesu podejmowania decyzji przez algorytm. Przedstawiony interpretacyjny system DL wykazuje potencjał jako narzędzie wspomagające podejmowanie decyzji w diagnostyce zmian chorobowych w wątrobie; jednakże wpływ kliniczny narzędzia wspomagającego podejmowanie decyzji musi być zweryfikowany w badaniu prospektywnym, zanim będzie można rozważyć jego włączenie do praktyki klinicznej.