

**Recenzja rozprawy doktorskiej mgr. inż. Marii Molik pt. „Analiza i prognozowanie wyników leczenia przewlekłej białaczki limfocytowej B-komórkowej z zastosowaniem metod bayesowskich”**

promotor rozprawy: prof. nzw. dr hab. inż. Piotr Ładyżyński

Recenzowana rozprawa doktorska dotyczy modelowania efektów leczenia przewlekłej białaczki limfocytowej CLL z zastosowaniem metod bayesowskich. Danymi badawczymi były zarówno wyniki zaczerpnięte z publikacji naukowych porównujących efekty różnych schematów leczenia, onkologicznych rejestrów w Stanach Zjednoczonych Ameryki oraz w Europie, jak również własne dane zgromadzone w systemie NetBIAL. Narzędziem rozwijanym w pracy były sieci bayesowskie, wsparte modelem ryzyka oraz analizą statystyczną. Badania mają wielotorowy charakter, podejmując zarówno analizę porównawczą różnych leków, różnicowanie grup chorych pod względem wieku i ogólnego stanu zdrowia, prognozowania dynamicznego stanu zdrowia oraz efektów wybranego schematu leczenia. W ramach autorskich narzędzi w pracy przedstawiony jest system NetBIAL, w którego tworzenie Autorka rozprawy miała istotny wkład, jak przede wszystkim opracowanie dynamicznych sieci bayesowskich *Stan zdrowia* oraz *Leczenie*. Wymienione treści mają na celu dowieść słuszności założonego celu rozprawy sformułowanego w postaci następujących czterech tez:

1. Sieciowa metaanaliza przeżywalności wykorzystująca wielomiany cząstkowe pierwszego rzędu do opisu parametrów reprezentujących efekt leczenia CLL pozwala na dokładniejsze dopasowanie modelu ryzyka do dostępnych danych niż dotychczas stosowana analiza oparta na modelach Weibulla i log-logistycznym.
2. W dotychczas przeprowadzonych badaniach klinicznych logarytm ryzyka dla czasu bez progresji CLL był odwrotnie proporcjonalny do czasu ( $\lambda$ ) a dla przeżycia całkowitego był proporcjonalny do kwadratu czasu.
3. Model zakładający zróżnicowany wpływ wieku i stadium zaawansowania CLL na prawdopodobieństwo remisji choroby dla każdego z testowanych dotychczas leków odzwierciedla wyniki badań klinicznych dokładniej niż model zakładający jednakowy wpływ tych czynników.
4. Jest możliwe opracowanie dynamicznej sieci bayesowskiej, która w skuteczny sposób pozwala przewidywać uśredniony wynik leczenia pacjentów z CLL w odniesieniu do całkowitego czasu przeżycia.

Charakterystyka zawartości pracy:

Recenzowana rozprawa składa się z 9 rozdziałów, 4 dodatków tabelarycznych i liczy 186 stron. Bibliografia zawiera 146 pozycji literaturowych oraz bazodanowych. Na wstępie zamieszczony jest wykaz skrótów stosowanych w tekście.

Pierwszy rozdział rozprawy zawiera krótki zarys problemu badawczego w aspekcie potencjału wspomagania leczenia przewlekłej białaczki limfocytowej, definiuje cel pracy i formułuje szczegółowe tezy. Przedstawiony jest w nim również przegląd kolejnych rozdziałów.

Drugi rozdział wprowadza czytelnika w zagadnienia medyczne związane z przewlekłą białaczką limfocytową. Przedstawiony jest w nim zarys klasyfikacji typów białaczek, różnicujących antygenów oraz stwierdzanych mutacji genów mających najważniejszy wpływ na agresywność i przebieg choroby. Rozdział przytacza z literatury stosowane czynniki prognostyczne, klasyfikacje zaawansowania choroby oraz obszernie dane statystyczne dotyczące częstości występowania poszczególnych grup chorych oraz czasów przeżycia. Ciąg dalszy rozdziału charakteryzuje procedury leczenia CLL oraz klasyfikację stosowanych leków.

Rozdział trzeci zawiera podstawy metaanalizy definiując modele z efektem stałym oraz zmiennym, wskaźniki metryczne, stosowany w pracy test jednorodności Q Cochra oraz zarys podstaw sieci bayesowskich.

W rozdziale czwartym przedstawiona została autorska metaanaliza bezpośrednia efektywności stosowanego leczenia CLL przeprowadzona na podstawie danych literaturowych. Dowodzi ona na wstępie zarówno obszerności badań w tym zakresie, jak też wykonanego pracochłonnego przeglądu literatury prowadzącego do selekcji możliwych do użycia w badaniu danych źródłowych. Przedstawione są metody bezpośredniej metaanalizy, w tym bayesowskie, oraz wyniki przeprowadzonych analiz wskaźników ilorazu szans, współczynnika ryzyka i różnicy ryzyka w kontekście porównania w parach różnych leków.

W rozdziale piątym przedstawiono metaanalizę sieciową efektywności leczenia CLL. Na podstawie publikacji prezentujących wyniki porównań efektów leczenia różnymi lekami (zwykle grupą referencyjną byli pacjenci leczeni chlorambucylem) skonstruowana została sieć powiązań badań klinicznych. Dla tych danych przeprowadzono najpierw analizę spójności wyników, a następnie metaanalizę sieciową w celu określenia najbardziej efektywnych schematów leczenia. W rozdziale zawarto również analizę wpływu wybranych czynników na efekt leczenia CLL.

W następnym rozdziale przedstawiono metaanalizę sieciową efektywności stosowanego leczenia CLL w odniesieniu do czasu przeżycia. Zdefiniowano i zastosowane w nim modele oparte na wielomianach cząstkowych, zbudowano stosowne sieci zależnie od ocenianego wskaźnika przeżycia oraz wykonano symulacje. W analizie oceniane były charakterystyki ryzyka (bezwzględne oraz względne), oraz odsetka pacjentów bez progresji CLL. Zaproponowany został model ryzyka oraz wykonano symulacje czasu całkowitego przeżycia.

W rozdziale siódmym zaprezentowano założenia systemu do gromadzenia danych pacjentów z CLL (NetBIAL) wprowadzonych przez ośrodki współpracujące. Przedstawiono zakresy danych występujących w formularzach internetowych oraz liczebności pozyskanych wyników badań.

Rozdział ósmy dotyczy projektowania dynamicznych sieci bayesowskich i określa zaproponowane przez Autorkę rozprawy struktury sieci oraz parametry poszczególnych węzłów. Zawarte są w nim

wyniki symulacji przeżycia w czasie w odniesieniu do baz SEER\*Stat oraz EUROCARE, jak również innych istotnych wskaźników chorobowych. Porównano również wyniki symulacji leczenia z danymi pochodzącymi z ośmiu publikacji naukowych oraz z danymi z bazy NetBIAL.

Ostatni rozdział rozprawy podsumowuje uzyskane wyniki badań oraz na koniec wymienia efekty pracy uważane przez Autorkę za najważniejsze.

Ocena rozprawy doktorskiej z uwzględnieniem wytycznych Rady Naukowej IBIB PAN:

Recenzowana rozprawa doktorska dotyczy metod komputerowych wspomagania procesu leczenia chorych na przewlekłą białaczkę limfocytową. Głównym celem pracy było stworzenie narzędzi oraz modeli symulujących rozwój choroby w czasie. Jest ona poparta licznymi danymi pochodzącymi z literatury, publicznych baz danych oraz systemu NetBIAL i ma charakter interdyscyplinarny, częściowo doświadczalny, z naciskiem na modelowanie procesów chorobowych. Postawione tezy pracy są ściśle związane z jej tematyką, czyli modelowaniem rozwoju choroby oraz wskaźników przeżywalności dla pacjentów ze zdiagnozowaną CLL. Pewne zastrzeżenia mogą budzić dwie pierwsze tezy, których sformułowanie sugeruje bardziej konkluzje z przeprowadzonych badań.

Rozprawa doktorska na różnych etapach wykorzystuje dane pochodzące z bardzo szerokiej i dokładnej analizy danych literatury światowej, jak również publicznych baz danych. Przyniesiony przegląd literatury oraz kwalifikacja danych do badań dowodzi bardzo rzetelnego podejścia do problemu badawczego. Autorka rozprawy przekonująco dowodzi przeprowadzonej selekcji danych, zaproponowanych metod badawczych oraz weryfikacji uzyskanych wyników symulacyjnych. Jako przykład może posłużyć selekcja 23 artykułów do metaanalizy bezpośredniej efektywności leczenia CLL (rozdział 4) spośród 8 308 pozycji otrzymanych w wyniku wyszukiwania hasła kluczowego „chronic lymphocytic leukemia” w bazie PubMed. Zastosowane przez Autorkę rozprawy kryteria dowodzą przyjęcia jako celu możliwie jak najwyższego poziomu wiarygodności statystycznej wyników badań.

Jako oryginalny dorobek i główny wkład w rozwój dyscypliny naukowej Autorka wymienia:

- metaanalizę bezpośrednią i sieciową oceny skuteczności leczenia CLL poprzez prawdopodobieństwo remisji choroby, w tym z uwzględnieniem płci oraz stopnia zaawansowania choroby,
- metaanalizę sieciową czasu całkowitego przeżycia oraz przeżycia bez progresji choroby z wyborem modeli tych parametrów,
- współudział w tworzeniu oraz wdrażaniu systemu NetBIAL,
- zaprojektowanie i implementacja dynamicznych sieci bayesowskich symulujących zmiany w czasie stanu zdrowia oraz wyniku leczenia CLL.

Należy stwierdzić, iż uzyskane wyniki badań są na wysokim światowym poziomie, a zaproponowane rozwiązania są autorskie dla tego problemu badawczego. Na szczególne wyróżnienie zasługują zaprojektowane dynamiczne sieci bayesowskie, których konstrukcja oraz parametryzacja przeprowadzona przez Autorkę dowodzi wysokiej znajomości zarówno narzędzi, jak też badanych procesów chorobowych. Zawarte w rozprawie wyniki są prawidłowo weryfikowane oraz wsparte

analizą statystyczną ich wiarygodności. Również dane źródłowe badań są na wielu etapach analizowane statystycznie i poparte cytowaniami.

Przedstawione w rozprawie autorskie rozwiązania problemów analizy danych oraz symulacji procesów chorobowych stanowią istotny wkład w praktyczne zastosowania techniki w medycynie. Zakres rozprawy najbardziej odpowiada dyscyplinie Biocybernetyka i Inżynieria Biomedyczna. Aspekty praktycznego zastosowania systemu NETBIAL oraz dynamicznych sieci do modelowania spodziewanego procesu chorobowego w zależności od kondycji pacjenta, wskaźników genetycznych oraz proponowanego schematy leczenia składają się na duży potencjał rozprawy.

Uwagi merytoryczne do tekstu rozprawy:

W procesie oceny rozprawy pojawiły się pewne wątpliwości co do poprawności niektórych sformułowań oraz wywodów. Większość z nich ma charakter edytorski, a jedynie nieliczne bardziej ogólne. Do uwag ogólnych zaliczam:

1. Przeprowadzona analiza spójności potwierdza część relacji wzajemnych różnych schematów leczenia. Może ona wzmacniać wiarygodność wniosków dla grup tworzących te relacje. Jednakże w pracy nie wykonano analizy efektów leczenia lekami tylko dla tak potwierdzonych relacji. Wykonując analizę sieciową dla wszystkich leków (czyli również tych bez weryfikacji spójności) otrzymano wskazanie na bendamustynę, co de facto opiera się na doniesieniach wyłącznie jednej pracy [95].
2. Czy istnieje możliwość oceny korelacji wyniku sieci *Leczenie w węźle przeżycie* z rzeczywistym przeżyciem pod względem zastosowanej terapii? Jeżeli tak, jaki odsetek pacjentów miał najdłuższe symulowane siecią przeżycie dla zastosowanej terapii w porównaniu z symulacją opartą na innych terapiach?
3. W pracy występują wielokrotne podsumowania – w rozdziałach oraz na końcu rozprawy, co skutkuje zbędnymi powtórzeniami. Podobnie we prowadzeniach do kolejnych rozdziałów wielokrotnie uzasadniany jest cel badawczy. W opinii recenzenta jest to zbędne. Ponadto w pracy występują wielokrotnie wywody oraz statystyki medyczne, które wielokrotnie nie mają przełożenia na opracowane autorskie metody komputerowe, natomiast brakuje szerszego tła dla stosowanych metod technicznych.

W opinii recenzenta wskazane by było odniesienie się do powyższych uwag w trakcie publicznej obrony doktoratu.

Do uwag natury edytorskiej zaliczam:

4. Co ma oznaczać maksymalizacja OD (36%) na stronie 36?
5. Wzór 4.1 – co oznacza indeks  $k-1$ ?
6. Rysunek 3.3 – brak oznaczenia prawdopodobieństw sieci,
7. Tekst i wzór 4.6 – raz jest  $\log it$ , innym razem *logit*.
8. Wzór 5.1 jest powtórzeniem wzoru 4.5.
9. Tabela 5.1, pierwsza kolumna – czy badany jest węzeł czy powiązanie? Dlaczego nie wszystkie relacje zostały zbadane, np. trójkąty CC-2CdA-CMC oraz CLB-FCLB-F?
10. Wzór 6.2 – brak znaku =.
11. Strona 65 – czy nie poprawnie jest „anglojęzycznych” zamiast „angielskojęzycznych”?

12. Wzór 6.3 i nast. – pewne wątpliwości mogą budzić oznaczenia, czy wszędzie chodzi o logarytm naturalny? W tekście logarytm ryzyka ( $h_{At}$ ) oznacza samo  $h_{At}$ , czy  $\ln(h_{At})$ ?
13. Strona 81, 1 linia – czy chodzi o Tab. 6.3 zamiast podanej 6.2?
14. Rysunek 6.6 – czy jest możliwe rozpoczynanie się przebiegu od wartości innej niż 100%, jak ma to miejsce w pracy?
15. Proszę o wyjaśnienie wartości w Tab. 8.1.
16. Czy (wzór 8.7) niezależnie od  $m$  transformacja jest zawsze bezpośrednio do stadium III,IV?
17. Jak przebiega wybór wartości  $x$  we wzorze 8.25?
18. Strona 123 – sformułowanie „na innym tempie” jest niepoprawne gramatycznie.
19. Rys. 8.18 – symulacja sieciowa zbiega do 0, w pracach [133,140] do wartości większej od 0. Wskazana jest dyskusja tego problemu.
20. Strona 125 – w tekście jest określenie czasu symulacji 25 lat, na rys. 8.20 jest 15 lat.
21. Strona 124 – na jakiej podstawie stwierdzono, iż dopiero po 15 latach wyniki symulowane zbliżają się do klinicznych?
22. Wzór 8.39 – Leczenie  $k$ , powinno być  $m$ . Czemu odpowiadają trzy kolejne wartości prawdopodobieństw?
23. Wzory 8.83 i 8.92 – jak są przyporządkowywane do węzła kolejne wartości prawdopodobieństw?

Ponadto w pracy można znaleźć błędy literowe (odmiany wyrazów, metaanaliza bayerowska s. 43 i 44), które powinny zostać wyeliminowane na etapie redakcji tekstu. Pewnym mankamentem jest liczba użytych skrótów, których czasem myślą się również Autorce rozprawy, np. strona 53.

Powyższe nieznaczące błędy edytorskie są na tyle marginalne, iż można stwierdzić, że praca została starannie napisana. Wymienione w recenzji uwagi są natury uzupełniającej i w żadnym stopniu nie negują wysokiego poziomu rozprawy, która spełnia wymagania na stopień w wyrażnym nadmiarze. Dowodzi temu zarówno zakres danych badawczych, złożoność problemu badawczego, jak również zaawansowany poziom zaproponowanych rozwiązań autorskich.

Konkludując, stwierdzam, iż recenzowana rozprawa doktorska Pani mgr. inż. Marii Molik pt. „Analiza i prognozowanie wyników leczenia przewlekłej białaczki limfocytowej B-komórkowej z zastosowaniem metod bayesowskich” w dyscyplinie Biocybernetyka i Inżynieria Biomedyczna spełnia wymagania wynikające z Ustawy o stopniach naukowych i tytułach naukowych oraz o stopniach i tytułach w zakresie sztuki z dn. 14 marca 2003 (Dz. U. Nr 65, poz. 595, z późn. zm.) i na tej podstawie wnoszę o dopuszczenie do publicznej obrony.

Z poważaniem



Tomasz Markiewicz