STRESZCZENIE ROZPRAWY DOKTORSKIEJ

mgr inż. Agnieszka Suchwałko

Zastosowanie analizy statystycznej do identyfikacji bakterii na podstawie widm dyfrakcyjnych kolonii bakterii

Autoreferat na podstawie publikacji

Promotor: Prof. n. tech. dr hab. n. fiz. inż. lek. med. Halina Podbielska
Promotor pomocniczy: dr inż. Igor Buzalewicz

Cel pracy:

Opracowanie szybkiej i dokładnej metody identyfikacji bakterii na podstawie analizy statystycznej cech liczbowych ekstrahowanych z widm dyfrakcyjnych Fresnela kolonii bakterii hodowanych na podłożach stałych oraz bazy danych widm dyfrakcyjnych kolonii bakterii, nadającej się do wdrożenia w praktyce mikrobiologicznej.

Tezy pracy:

1. Uwzględnienie w procesie identyfikacji naturalnego podziału widm dyfrakcyjnych na obszary oraz średnicy widm ma istotny wpływ na możliwości identyfikacji bakterii.
2. Wyznaczenie na podstawie wartości pikseli właściwych cech liczbowych, uwzględniających charakterystyczną teksturę i morfologię zarejestrowanych widm dyfrakcyjnych Fresnela kolonii bakterii, zapewnia optymalne wykorzystanie informacji przez nie niesionej.
3. Wybór odpowiednich cech liczbowych spośród cech ekstrahowanych z widm dyfrakcyjnych Fresnela kolonii bakterii pozwala na budowę optymalnych modeli klasyfikacyjnych.
4. Zastosowanie normalizacji zarejestrowanych widm dyfrakcyjnych Fresnela kolonii bakterii zapewnia ich dokładniejsze porównanie między sobą, co prowadzi do zmniejszenia wartości błędu identyfikacji.
5. Metoda cechuje się dokładnością, która umożliwia identyfikację na poziomie serowarów[1] bakteryjnych.
6. Wprowadzenie nieznacznych modyfikacji układu optycznego zapewnia powtarzalność pomiarów niezależnie od osoby wykonującej rejestrację z zaledwie nieznacznym pogorszeniem otrzymywanych wyników.
7. Zastosowanie właściwych sposobów i algorytmów klasyfikacji[2] i identyfikacji[3] oraz weryfikacji[4] otrzymanych wyników pozwala na uzyskanie rezultatu identyfikacji bakterii z najmniejszym możliwym błędem.

---

[1] Serowar (serotyp) – odmiana w obrębie gatunku bakterii wyłoniona za pomocą metod serologicznych, wykazujących różnice w budowie antygenowej.

[2] Klasyfikacja – grupowanie w tym konkretnym przypadku widm dyfrakcyjnych kolonii bakterii na podstawie cech wspólnych, którymi są ekstrahowane z widm cechy liczbowe. Klasyfikacja obejmuje cały proces grupowania (od wyboru cech poprzez budowę modeli po przypisywanie do konkretnej grupy) odbywający się z wykorzystaniem algorytmów QDA i SVM.

[3] Identyfikacja – ostateczny etap klasyfikacji przyporządkowujący dane widmo dyfrakcyjne do jednej z grup.

[4] Weryfikacja – (ocena jakości) procedura mająca na celu stwierdzenie czy zbudowany model klasyfikacyjny poprawnie grupuje widma dyfrakcyjne. Weryfikacja odbywa się metodą CV z zastosowanym algorytmem losowania warstwowego (ang. stratified sampling), który zapewnia homogeniczny rozkład widm w zbiorach uczącym i testowym.

Przeprowadzone prace badawcze i uzyskane wyniki:

Uzyskane wyniki prac badawczych prowadzonych w ramach doktoratu zostały przedstawione w 6 publikacjach naukowych (jeden rozdział w książce anglojęzycznej, 2 prace z Listy Filadelfijskiej, 3 publikacje w materiałach konferencyjnych), jedno krajowe zgłoszenie patentowe.

Pierwsze badania, których wyniki znalazły się w pracy [1], miały na celu pokazanie, iż dane ekstrahowane z widm dyfrakcyjnych Fresnela kolonii bakterii zarejestrowanych przy pomocy specjalnie zaprojektowanego układu optycznego umożliwiają identyfikację bakterii z dużą dokładnością. Zaproponowany został podział widm na obszary, z których następnie wydobyte zostały cechy liczbowe. Podział na obszary wynika z naturalnego przestrzennego rozkładu widm na obszary (zbliżone kształtem do pierścieni kołowych), które znacząco różnią się między grupami bakterii. Jako cechy liczbowe opisujące analizowane widma dyfrakcyjne wykorzystane zostały wartość średnia oraz odchylenie standardowe wartości pikseli[5] będące miarami jasności i szorstkości poszczególnych obszarów widm. Przeprowadzona została eksploracyjna analizy danych w celu zobrazowania potencjału danych w zakresie różnicowania grup bakterii. Następnie wybrano najlepsze cechy liczbowe za pomocą analizy wariancji ANOVA i na ich podstawie zbudowano model klasyfikacyjny LDA (Linear Discriminant Analysis – Liniowa Analiza Dyskryminacyjna). **Weryfikacja otrzymanych rezultatów za pomocą walidacji krzyżowej wykazała błąd identyfikacji bakterii na poziomie 5,97%.**

Kolejne prowadzone prace zapoczątkowało powiększenie bazy danych widm dyfrakcyjnych kolonii bakterii [2]. Analiza nowego, znacznie większego zbioru danych była konieczna do potwierdzenia możliwości identyfikacji bakterii na podstawie jedynie prostych cech ekstrahowanych z widm kolonii bakterii. Uzyskanie wyniku na poziomie 8,11% błędu identyfikacji dla nowej, niemal dwukrotnie większej bazy danych, pozwoliło przypuszczać, iż wstępne przygotowanie danych połączone z zastosowaniem bardziej zaawansowanego aparatu statystycznej analizy danych oraz ekstrakcja i wybór lepiej różnicujących grupy bakteryjne cech liczbowych, znacząco poprawi otrzymane wyniki.

Następne eksperymenty, opisane w pracy [3], oprócz dotychczas zastosowanych technik przetwarzania i analizy danych zostały uzupełnione o normalizację tła widm, która miała na celu ujednolicenie skali szarości obrazów cyfrowych widm dyfrakcyjnych kolonii bakterii, przetestowanie działania klasyfikatorów: QDA (Quadratic Discriminant Analysis – Kwadratowa Analiza Dyskryminacyjna) i SVM (Support Vector Machine – Maszyna Wektorów Nośnych) oraz wyznaczenie czułości i specyficzności w wersji dla wielu klas. Zastosowanie normalizacji tła oraz klasyfikatora QDA przy podziale widm na 10 obszarów pozwoliło na **zmniejszenie błędu identyfikacji do wartości 3,14% przy czułości 100% i specyficzności 96,35%.**

Następnym krokiem usprawniania metody identyfikacji bakterii, zaprezentowanym na konferencji SPIE BIOS w Monachium [4], było przetestowanie morfologicznych i teksturalnych cech liczbowych, opartych na momentach statystycznych oraz porównanie wpływu na wyniki analizy wariancji ANOVA z dywergencją Fishera (SNR – Signal to Noise Ratio). Dodatkowo, niezbędne było wprowadzenie dwukrokowego wyboru najlepiej separujących bakterie cech: najpierw zestawów cech bez podziału na obszary (np. średnia, odchylenie standardowe) najlepiej separujących grupy bakterii, a następnie najlepiej separujących grupy bakterii cech spośród wybranych wcześniej zestawów (np. średnia w obszarach 1 i 3 licząc od środka widma). **Najskuteczniejszym klasyfikatorem okazał się QDA, dając błąd 0,857% w połączeniu z rankingiem cech metodą SNR dla wybranych 18 najlepiej różnicujących cech. Uzyskana dla najlepszego wyniku czułość i specyficzność wieloklasowa są bliskie 100%.**

Podsumowanie dotychczasowych wyników oraz porównanie z konkurencyjną metodą identyfikacji bakterii amerykańskiej grupy badaczy (wykorzystującej inny układ optyczny oraz analizę) zawarte zostało w rozdziale książki o zasięgu międzynarodowym [5]. Badania amerykańskiej grupy różnią się od prac prowadzonych na Politechnice Wrocławskiej nie tylko sposobem zbierania i analizy

[5] Pojęcie wartości piksela odnosi się do jednego z możliwych poziomów dostępnej skali. Widma rejestrowane były jako obrazy 8-bitowe czyli w 256 poziomowej skali szarości. Oznacza to możliwe wartości z przedziału [0-255], gdzie 0 to czerń, a 255 to biel.

danych, ale także prezentacją wyników oraz wielkością zbiorów danych poddanych analizie. Nasze dane są większe i bardziej różnorodne. Prowadzone w obu grupach badania zakładają również inne warunki hodowli, więc jednoznaczne porównanie nie było możliwe. Jednak podejmując próbę zestawienia opublikowanych wyników przez obie grupy dla największych zbiorów danych, co dla Amerykanów oznacza zbiór 4 szczepów, a dla nas zbiór obejmujący 7 grup bakterii, **większa wartość czułości oraz mniejszy o ponad 1% błąd identyfikacji przemawiają na korzyść naszej metody.**

Ostatni etap badań podjętych w ramach doktoratu został podsumowany w pracy z Listy Filadelfijskiej [6]. Praca dotyczy optymalizacji i standaryzacji całej metody, zarówno sposobu przygotowania próbek, układu optycznego, warunków rejestracji widm dyfrakcyjnych, jak i analizy danych. Metoda wymagała doprecyzowania procedur w związku z bezwzględnym wymogiem zapewnienia powtarzalności eksperymentów, niezależnie osób je prowadzących. Kolejne modyfikacje metody polegały na przetestowaniu możliwości identyfikacji bakterii zarejestrowanych dla 7 różnych odległości próbki (szalki Petriego z wyhodowanymi koloniami) od obiektywu kamery rejestrującej widma dyfrakcyjne i wyborze optymalnej odległości. Ponadto, testowano różne czasy inkubacji, potrzebne do wyhodowania kolonii o średnicy najbardziej optymalnej do dalszych pomiarów. Dołożono również nową cechę opisującą promień widma dyfrakcyjnego kolonii bakterii, która dobrze różnicuje bakterie, ponieważ jest uzależniona od stadium rozwoju kolonii, a to znacznie różni się pomiędzy grupami bakterii. Zastosowano też metodę losowania warstwowego (ang. stratified sampling), celem zwiększenia reprezentatywności prób do walidacji krzyżowej. Porównane zostały również obie ścieżki eksperymentalne: pierwotna i zoptymalizowana. **Zmodyfikowana metoda okazała się skuteczniejsza niż jej poprzednia wersja, osiągając błąd identyfikacji na poziomie 1,34% oraz czułość i specyficzność wieloklasową odpowiednio 97,59% i 99,03%.** Optymalny okazał się klasyfikator SVM.

Wnioski:

W trakcie prowadzonych badań metoda identyfikacji bakterii na podstawie analizy widm dyfrakcyjnych Fresnela kolonii bakterii została poddana wielu modyfikacjom, które znacząco wpłynęły zarówno na jej dokładność, jak i przydatność do zastosowania w praktyce. Z metody czysto badawczej została przekształcona w technikę spełniającą wymogi techniki stosowanej komercyjnie. Naturalnie wachlarz możliwych modyfikacji i usprawnień nie został w pełni wyczerpany, ale uzyskane dzięki zastosowanym zmianom rezultaty są w pełni satysfakcjonujące.

Metoda wykorzystuje unikalne właściwości światła rozproszonego na kolonii bakterii w postaci widma dyfrakcyjnego Fresnela. Zastosowanie odpowiednio dobranego aparatu statystycznej analizy danych pozwoliło wykorzystać w sposób optymalny informację zarejestrowaną w postaci widm dyfrakcyjnych kolonii bakterii w celu ich różnicowania. Metoda nie wymaga stosowania żadnych specjalistycznych odczynników, a jedynie standardowych materiałów (podłoży hodowlanych) znajdujących się w laboratorium mikrobiologicznym, co do których istnieje wyłącznie wymóg dobrej przejrzystości (transparentności). Cały proces identyfikacji bakterii jest przeprowadzany automatycznie, a udział personelu badawczego ogranicza się do przygotowania próbek. Takie wymagania zapewniają niskie koszty działania i brak drogich szkoleń personelu. Czas działania jest ograniczony do czasu niezbędnego na wyhodowanie kolonii bakterii o średnicy umożliwiającej rejestrację widma dyfrakcyjnego. Na przeprowadzenie analizy potrzeba kilka minut, co jest konkurencyjne w stosunku do większości metod z wyjątkiem testów (np. immunoenzymatycznych). Za pomocą opracowanej metody można identyfikować bakterie, których widma wcześniej zostały zapisane w bazie referencyjnych widm dyfrakcyjnych. Nowe (np. zmutowane) gatunki czy szczepy bakterii będą mogły zostać dodane do bazy danych i niezwłocznie podlegać identyfikacji bez konieczności ingerencji w samą metodę. Metoda jest w stanie identyfikować również mieszaniny bakterii, ponieważ przy odpowiednim rozcieńczeniu, każda kolonia wyrasta z pojedynczej komórki bakteryjnej. Jeżeli dana kolonia nie styka się z innymi na płytce, podlega identyfikacji niezależnie od pozostałych kolonii znajdujących się w próbce. Takie

podejście pozwala również, na określenie procentowej zawartości w badanej próbce określonych grup bakterii. Wykazana została również skuteczność metody w przypadku identyfikacji serowarów bakteryjnych.

Duży potencjał aplikacyjny opracowanej metody został dostrzeżony również przez podmioty gospodarcze. We współpracy z firmą Bioavlee Sp. z o. o. prowadzone były na Politechnice Wrocławskiej prace badawczo-rozwojowe prezentowanej metody identyfikacji bakterii. Obecnie firma ta zajmuje się wdrożeniem komercyjnego urządzenia opierającego swoje działanie na koncepcjach opracowanych przez Grupę Bio-Optyki z Katedry Inżynierii Biomedycznej, w której prowadzone były badania.

Prace stanowiące podstawę ubiegania się o stopień doktora:

1. **Suchwalko A**, Buzalewicz I, Podbielska H (2012) *Computer-based classification of bacteria species by analysis of their colonies Fresnel diffraction patterns.* In: Miller BL, Fauchet PM, editors. Proceedings of SPIE. Vol. 11. p. 82120R – 82120R – 13. http://www.opticsinfobase.org/abstract.cfm?URI=BIOMED-2012-BSu5A.5

2. Podbielska H, Buzalewicz I, **Suchwalko A**, Wieliczko A (2012) *Bacteria Classification by Means of the Statistical Analysis of Fresnel Diffraction Patterns of Bacteria Colonies.* Biomedical Optics and 3-D Imaging. Washington, D.C.: OSA. p. BSu5A.5. http://www.opticsinfobase.org/abstract.cfm?URI=BIOMED-2012-BSu5A.5

3. **Suchwalko A**, Buzalewicz I, Wieliczko A, Podbielska H (2013) *Bacteria species identification by the statistical analysis of bacterial colonies Fresnel patterns.* Opt Express 21: 11322–11337, **IF: 3.525, Punkt MNiSW (2013): 45** http://www.opticsinfobase.org/abstract.cfm?URI=oe-21-9-11322

4. **Suchwalko A**, Buzalewicz I, Podbielska H (2013) *Identification of bacteria species by using morphological and textural properties of bacterial colonies diffraction patterns.* In: Remondino F, Shortis MR, Beyerer J, Puente León F, editors. Proceedings of SPIE. pp. 87911M – 1–87911M – 7. http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1691637

5. **Suchwalko A**, Buzalewicz I, Podbielska H (2013) *Statistical identification of bacteria species*. In: Méndez-Vilas A, editor. Microbial pathogens and strategies for combating them: science, technology and education. Badajoz, Spain: Formatex Research Center. pp. 711–721. http://www.formatex.info/microbiology4/vol1/711-721.pdf

6. **Suchwałko A**, Buzalewicz I, Podbielska H (2014) *Bacteria identification in an optical system with optimized diffraction pattern registration condition supported by enhanced statistical analysis*. Opt Express 22: 26312–26327. **IF: 3.488, Punkt MNiSW (2014): 45** http://www.ncbi.nlm.nih.gov/pubmed/25401664.

7. **A. Suchwałko**, H. Podbielska, I. Buzalewicz, „Sposób identyfikacji gatunku bakterii" , nr zgłoszenia P 400116 z dn. 24.07.2012