



Politechnika Wroclawska



Wydział Elektroniki,
Fotoniki i Mikrosystemów

Wrocław, 15 stycznia 2024

dr hab. inż. Grzegorz Świrniak., prof. PWr
Katedra Metrologii Elektronicznej i Fotonicznej
Wydział Elektroniki, Fotoniki i Mikrosystemów
Politechnika Wroclawska
ul. B. Prusa 53/55, 50-317 Wrocław
e-mail: grzegorz.swirniak@pwr.edu.pl
tel. +48 71 320 62 32

RECENZJA

**rozprawy doktorskiej mgra inż. Krzysztofa Smykały
pt.: „Zastosowanie metod uczenia maszynowego i pomiarów hiperspektralnych
w diagnostyce uprawy *Solanum lycopersicum*”**

Promotor rozprawy: dr hab. inż. Michał Tomaszewski, prof. uczelni
Promotor pomocniczy: dr inż. Bogdan Ruszczak

Podstawa formalna opracowania recenzji

Recenzję wykonano w odpowiedzi na uchwałę Rady Naukowej Dyscypliny Automatyka, Elektronika, Elektrotechnika i Technologie Kosmiczne z 8 listopada 2023 r., zgodnie z zapisami Ustawy Prawo o szkolnictwie wyższym i nauce z 20 lipca 2018 r.

1. Rozwiązywany problem naukowy

Tematyka przedstawionej do recenzji pracy doktorskiej została jasno sformułowana we wstępie manuskryptu i dotyczy bezkontaktowej, nieinwazyjnej oceny wybranych jednostek chorobowych pomidora zwyczajnego (*Solanum lycopersicum*) na podstawie pomiarów hiperspektralnych promieniowania odbitego od roślin i przy wykorzystaniu algorytmów uczenia maszynowego. Motywacją autora było zastąpienie metod tradycyjnych (czyli takich, które bazują na wiedzy i doświadczeniu eksperta bądź metod laboratoryjnych) metodami teledetekcyjnymi. Prace badawcze w tym obszarze mają duże znaczenie praktyczne, ponieważ umożliwiają wczesną ocenę zmian chorobowych w sposób zautomatyzowany i na dużej powierzchni uprawowej przy wykorzystaniu technik lotniczych (np. bezzałogowych statków powietrznych).

Rozwiązywanym problemem naukowym było opracowanie modelu uczenia maszynowego, który jest zdolny do rozpoznawania charakterystycznych wzorców w danych hiperspektralnych i do przewidywania obecności lub braku chorób roślinnych pomidora zwyczajnego. Tak postawiony problem ma charakter nowatorski, ponieważ, jak dotąd, nie przeprowadzono uporządkowanych i systematycznych analiz w tym obszarze. Jest on właściwy dla pracy doktorskiej w dziedzinie **nauk inżynieryjno-technicznych**, w dyscyplinie **automatyka, elektronika, elektrotechnika i technologie kosmiczne**. Wynika to z faktu, że rozwiązywany problem naukowy i przyjęta metodyka dotyczą elektronicznych systemów pomiarowych, technik i technologii pomiarowych, algorytmicznego przetwarzania danych oraz automatyzacji procesu podejmowania decyzji.

Cel pracy (str. 9) sformułowano jako stworzenie metody pozwalającej na wstępną, automatyczną diagnostykę obiektu za pomocą danych hiperspektralnych oraz algorytmów uczenia maszynowego na przykładzie diagnostyki wybranych chorób w uprawach roślin.

Zakres przeprowadzonych badań podzielono na kilka głównych etapów. Należały do nich (str. 9): (i) przegląd wybranych chorób bakteryjnych, wirusowych i grzybowych pomidorów, (ii) analiza obecnego stanu wiedzy na temat zastosowań metody wielospektralnej w diagnostyce roślin, (iii) przegląd dotychczasowych osiągnięć w zakresie diagnostyki chorób pomidora na podstawie danych wielospektralnych, (iv) opracowanie metody pomiarowej, (v) realizacja eksperymentu pomiarowego, mającego na celu pozyskanie danych hiperspektralnych w warunkach laboratoryjnych, (vi) wybór algorytmów uczenia maszynowego, (vii) opracowanie metody przetwarzania danych hiperspektralnych oraz (viii) przeprowadzenie analiz mających na celu ocenę skuteczności klasyfikacji wybranych chorób pomidora przy użyciu opracowanych algorytmów.

Teza badawcza (str. 10) brzmi następująco: Istnieje możliwość wykorzystania pomiarów hiperspektralnych refleksyjności z zakresu fal o długości 350-2500 nm i algorytmów uczenia maszynowego w procesie automatycznej diagnostyki *Solanum lycopersicum* dla potrzeb klasyfikacji wybranych jednostek chorobowych.

Tezę badawczą uzupełniają trzy tezy szczegółowe: i) Możliwe jest określenie algorytmów klasyfikacyjnych, które z większą efektywnością pozwalają na rozpoznanie następujących chorób pomidora zwyczajnego: alternarioza, antraknoza, bakteryjna cętkowatość pomidora, septorioza, zaraza ziemniaka w pomidorze z zastosowaniem ograniczonego zbioru danych uczących; ii) Możliwe jest określenie specyficznych długości zakresu widma odpowiadających określonym jednostkom chorobowym *Solanum lycopersicum*, iii) Hiperspektralne pomiary refleksyjności pozwalają na klasyfikację jednostek chorobowych *Solanum lycopersicum* w czasie szybszym niż metoda oceny wizualnej stosowana w praktyce.

W warstwie literalnej główna teza badawcza oraz tezy szczegółowe sformułowane są prawidłowo i dobrze nakreślają kierunek badań. W tezie szczegółowej i) brakuje jednak punktu odniesienia dla sformułowania „z większą efektywnością”. Z kolei teza szczegółowa iii) nie ma uzasadnienia w badaniach zrealizowanych przez Doktoranta, które nie dotyczą oceny wiedzy i umiejętności eksperta.

2. Ogólna charakterystyka rozprawy

Praca doktorska mgr inż. Krzysztofa Smykały ma charakter badań stosowanych, podejmowanych w celu zdobycia nowej wiedzy i zorientowanych na zastosowanie w praktyce. Zawiera badania eksperymentalne z wykorzystaniem aparatury optycznej i elektronicznej, z zastosowaniem algorytmów uczenia maszynowego w zadaniach identyfikacji i klasyfikacji oraz elementy analizy danych pomiarowych.

Rozprawa składa się z 18 wyróżnionych części, w tym z 14 rozdziałów numerowanych, i liczy 150 stron tekstu. Układ pracy jest poprawny i zawiera zarówno elementy potrzebne do zrozumienia tematyki badań jak i również niezbędne wyniki prac empirycznych i numerycznych. We wprowadzeniu Autor przedstawia w sposób ogólny tematykę pracy. Cel, zakres, teza oraz układ pracy są treścią rozdziału 2. Rozdział 3 przedstawia podstawowe informacje na temat pięciu chorób pomidorów (tj., alternarioza, antraknoza, bakteryjna cętkowatość, septorioza oraz tzw. zaraza ziemniaczana), które są brane pod uwagę w dalszych poczynaniach badawczych, a także opisuje podstawowe obszary zastosowań metrologii w diagnostyce chorób roślin. Rozdział 4 to przegląd literatury na temat zastosowania technik wielospektralnych w diagnostyce chorób pomidorów. Analiza ta pozwala Autorowi na sformułowanie konkluzji, według której dotychczasowe prace wymagają uzupełnienia przede wszystkim w zakresie różnorodności wykrywanych chorób, trafności klasyfikacji i częstotliwości pomiarów. W rozdziale 5 przedstawiono zarys metody diagnostycznej bazującej na spektralnych sygnaturach roślin, zarówno na obecnym etapie badań Autora jak i również perspektywy jej rozwoju. Rozdział 6 to opis metodologii eksperymentu pomiarowego, mającego na celu pomiary hiperspektralne roślin hodowanych w kontrolowanych warunkach (w komorach vegetacyjnych), dokonywane w różnych fazach wzrostu i rozwoju. Analiza dotyczyła dwóch odmian pomidora – *Benito* i *Polfast*, łącznie 36 roślin, zarażonych pięcioma odmiennymi patogenami, wywołującymi wymienione powyżej choroby. Zbiór kontrolny bez stwierdzonych chorób liczył 6 roślin. Rozdziały 7–9 prezentują główne elementy oryginalnego osiągnięcia Autora. Rozdział 7 zawiera opis klasyfikatorów, które zostały przetestowane przez Autora pod kątem przydatności w rozwiązaniu problemu klasyfikacji binarnej – czyli odpowiedzi na pytanie, czy badana roślina jest zarażona czy zdrowa, a także miary oceny klasyfikatorów. Rozdział 8 opisuje metodologię postępowania z danymi pomiarowymi na rzecz utworzenia zbioru treningowego oraz ocenę tych danych przy użyciu znormalizowanego wskaźnika vegetacyjnego (NDVI), opisującego kondycję roślinności. Rozdział 9 obejmuje analizy mające na celu znalezienie najodpowiedniejszego klasyfikatora binarnego, przeprowadzone dla 6 wybranych wariantów eksperymentu (grup roślin zakażonych 6 wariantami patogenów). Rozdział ten kończy się konkluzją mającą uzasadnić tezę badawczą pracy. W rozdziale 9 Autor przedstawia analizy mające na celu wytworzenie ścieżki postępowania z danymi hiperspektralnymi, a szczególności wskazanie pasma widmowego, które zawiera informacje istotne z punktu widzenia podejmowanego problemu naukowego. Rozważany jest także problem eliminacji cech wynikający z nadmiarowości informacji w danych hiperspektralnych. Wnioski płynące z analiz pozwalają Autorowi na potwierdzenie pierwszej i drugiej tezy szczegółowej. Rozdział 9 kończy opis prób klasyfikacji wieloklasowej.

Rozdział 10 prezentuje analizy działania zbudowanych klasyfikatorów w ujęciu czasowym, dla 6 wariantów eksperymentu (6 grup badawczych), co umożliwia ocenę przydatności metody hiperspektralnej do wczesnego wykrywania objawów chorobowych. Rozdział ten kończy się konkluzją mająca uzasadnić trzecią tezę szczegółową. Rozprawę kończą: podsumowanie, opis kierunków dalszych prac, bibliografia, streszczenie, spisy rysunków i tabel oraz załączniki.

3. Ocena merytoryczna rozprawy

Analiza stanu wiedzy

Autor dokonał ogólnego rozeznania literaturowego w wielu obszarach powiązanych tematycznie z dziedziną pracy, tj. wybranych chorób pomidorów, zastosowania technik wielospektralnych w diagnostyce chorób pomidorów i opis klasyfikatorów, które zostały przetestowane przez Autora pod kątem przydatności w rozwiązaniu problemu klasyfikacji binarnej. Wykaz literatury zamieszczony w dysertacji zawiera 103 pozycje w postaci książek, artykułów naukowych i źródeł internetowych, w większości wydane w ostatnim dziesięcioleciu, co świadczy o aktualnym stanie wiedzy, którą posługuje się Autor. Moje zastrzeżenie budzi brak odniesienia do literatury wyników symulacyjnych zawartych w rozdziałach 9.2 (analiza algorytmów klasyfikacyjnych) i 10 (analizy działania klasyfikatorów w ujęciu czasowym).

Metodyka pracy

Metodyka pracy jest prawidłowa. Obejmuje identyfikację i opis problemu badawczego, elementy rozeznania literaturowego i uporządkowania wiedzy w obszarach związanych z tematyką pracy badawczej, przeprowadzenie badań w warunkach laboratoryjnych, analizy mające na celu zbudowanie i selekcję modeli numerycznych, prace symulacyjne, krytyczną analizę wyników ich upowszechnienie, oraz końcowe wnioski

Na uznanie zasługuje umiejętność poruszania się Autora w wielu obszarach nauki i techniki, których znajomość i wykorzystanie były konieczne do właściwej realizacji pracy, a przede wszystkim w zakresie metod i procedur przetwarzania hiperspektralnych danych pomiarowych. Wysoko oceniam również znajomość i prawidłowe wykorzystanie przez Autora metod uczenia maszynowego, zwłaszcza w kontekście wyciągniętych na ich podstawie wniosków. Kolejnym wartościowym elementem pracy jest przeprowadzenie eksperymentów w warunkach laboratoryjnych, chociaż udział Autora w tych pracach wymaga doprecyzowania ze względu na wielodyscyplinowy i wieloautorski charakter prac.

Oryginalność rozwiązania problemu naukowego

Postawiony problem naukowy – opracowanie modelu uczenia maszynowego, który jest zdolny do rozpoznawania charakterystycznych wzorców w danych hiperspektralnych i do przewidywania obecności lub braku chorób roślinnych pomidora zwyczajnego, jest prawidłowo zidentyfikowanym zadaniem w obszarze dyscypliny naukowej *automatyka, elektronika, elektrotechnika i technologie kosmiczne*. Trafność wyboru tej problematyki wynika przede wszystkim z jej aktualności, złożoności rozwiązywanego problemu merytorycznego

oraz praktycznego znaczenia rozwiązywanego problemu teledetekcyjnej diagnostyce w upraw roślin.

Oryginalność naukową recenzowanej pracy cechuje zarówno jakość uzyskanych wyników, które mają charakter poznawczy i użyteczny, jak i metody i narzędzia badawcze, którymi Autor się posługiwał. **Do oryginalnych cech recenzowanej pracy zaliczam:**

- i) zaplanowanie i zrealizowanie dużej liczby eksperymentów laboratoryjnych;**
- ii) wskazanie pasm widmowych w obszarze SWIR (1800–1890 nm) oraz 2269 nm specyficznych dla wybranych jednostek chorobowych pomidora zwyczajnego;**
- iii) opracowanie strategii przetwarzania hiperspektralnych danych pomiarowych umożliwiających realizację sformułowanego problemu naukowego;**
- iv) systematyczną analizę danych hiperspektralnych przy użyciu wybranych algorytmów uczenia maszynowego.**

Oryginalność wkładu w dyscyplinę *automatyka, elektronika, elektrotechnika i technologie kosmiczne* dokumentują publikacje współautorstwa Pana mgr inż. Krzysztofa Smykały zamieszczone w bibliografii. Są to: **2** artykuły w międzynarodowych periodykach z listy JCR (*Journal of Ambient Intelligence and Smart Environments* IF = 2.105, *Scientific Reports* IF = 4.6) oraz **1** artykuł w innym czasopiśmie międzynarodowym, w którym jest pierwszym autorem.

Redakcja rozprawy

Praca napisana jest ładnym językiem polskim na dobrym poziomie. Treść jest komunikatywna i wystarczająco precyzyjna. Obecne są drobne błędy językowe i edytorskie, takie jak:

- między wartościami i ich jednostkami brakuje czasami spacji; edytując tekst, jednostki miar należy oddzielać od ich wartości;
- w pracach naukowych należy stosować nie tylko jednostki SI, ale i standardowy ich zapis, np. $m \cdot s^{-1}$ zamiast m/s;
- tekst powinien być sformatowany z zastosowaniem konsekwentnych reguł wyodrębniania akapitów;
- rysunki powinny być umieszczone na stronie pierwszego powołania się na nie, lub na stronie sąsiedniej, co nie zawsze ma miejsce;
- niefortunnym błędem językowym jest wyrażenie „zbieranie danych pomiarowych”; lepiej dane pozyskiwać, dokonywać ich akwizycji, analizować, przetwarzać, itd. Innym socjolektem jest „czysty zbiór danych” (Rozdz. 8.1);
- brakuje opisów osi na Rys. 51, 52;
- opisy rysunków i nagłówek tabel powinny być sporządzone w języku polskim, co nie zawsze ma miejsce.

Pozostałe błędy językowe nie zasługują na to, aby wymieniać je w zasadniczej części recenzji, dlatego zaznaczono je w tekście manuskryptu do wglądu Autora. Dużym walorem pracy są czytelne wykresy i tabele, znakomicie ilustrujące opisywane zagadnienia. Wartościowe są także podsumowania kończące główne rozdziały, które pomagają w porządkowaniu wiedzy.

4. Uwagi krytyczne, zagadnienia do dyskusji

Dysertacja doktorska mgr inż. Krzysztofa Smykały wymaga kilku dodatkowych komentarzy. Poniżej zamieszczam najważniejsze z nich.

- 1) Najważniejszym niedociągnięciem rozprawy jest brak jasno sprecyzowanego wkładu Autora w przedstawione wyniki. Po pierwsze, Autor zaznacza w tekście manuskryptu, że wykorzystał dane pochodzące z badań przeprowadzonych przez firmę QZ Solutions (Rozdz. 6). Jaki jest wkład Autora w prace laboratoryjne, pomiary, analizę danych, symulacje i proces wyciągania głównych wniosków zrealizowanych w tej firmie? Ponadto, w dysertacji powinien być przedstawiony udział merytoryczny Autora w powstaniu publikacji wieloautorskich, zamieszczonych w bibliografii. Proszę o uzupełnienie tych informacji.
- 2) Czy ortofotomapy z Rys. 18–20 ilustrują zmiany chorobowe uprawy pomidora? W tekście manuskryptu nie znalazłem informacji na ten temat.
- 3) Proszę o uzasadnienie, dlaczego analizy statystyczne w postępowaniu ze zbiorami danych (Rozdz. 8) dotyczyły takiego parametru rozkładu, jak mediana, a nie wartość średnia?
- 4) Pomiary reflektancji zostały wykonane z wysokości 5 cm „jako najmniej zaszumione obrazem tła, a jednocześnie najbardziej wiarygodne w procesie badawczym (...)” (Str. 54). Jak należy rozumieć stwierdzenie „najbardziej wiarygodne [pomiary] w procesie badawczym”? Czy 5 cm to rzeczywiście optymalna odległość w kontekście koncepcji teledetekcyjnej diagnostyki chorób roślin opisanej w rozdziale 5?
- 5) Ile ostatecznie sekwencji użyto do uczenia klasyfikatorów opisanych w Rozdz. 9?
- 6) Na podstawie przeprowadzonych analiz Autor stwierdził, że (Str. 126) „wyniki wskazują na możliwość diagnostyki badanych chorób w okresie krótszym niż 7 dni od zakażenia, co potwierdza trzecią tezę niniejszej rozprawy doktorskiej”. W kontekście tego stwierdzenia, proszę o ustosunkowanie się do następujących zagadnień:
 - i) W praktyce szybkość rozwoju chorób (i ich objawów) zależy m.in. od warunków wzrostu rośliny, w tym temperatury, wilgotności powietrza i podłoża, zagęszczenia roślin i cech gatunkowych. Rodzi się więc pytanie, czy wnioski z analiz przeprowadzonych w reżimie laboratoryjnym mogą być łatwo uogólnione na badania wykonywane w rzeczywistych warunkach?
 - ii) Autor w opisie eksperymentu pomiarowego (Str. 34) pisze, że „Pierwsze, widoczne dla człowieka objawy zarażenia, widoczne były po 3-5 dniach od zakażenia.” W świetle tego stwierdzenia, przytoczone powyżej ustalenia badań modelowych (Str. 126) nie potwierdzają więc trzeciej tezy pomocniczej (Str. 11). Proszę o komentarz.

- iii) Zmiany chorobowe obserwowane na młodych sadzonkach (niezwykle rzadko występujące w praktyce) nie postępują tak szybko jak zmiany obserwowane na roślinach dojrzałych, owocujących. Wynika to chociażby z porównania zdjęć młodych roślin zakażonych patogenami (Rys. 27, rozwój choroby w ciągu 14 dni od inokulacji) oraz ortofotomap w pełni wykształconych roślin (Rys. 18–20, intensywny rozwój choroby w ciągu 7 dni). Czy zatem wyniki klasyfikacji chorobowej opartej na danych hiperspektralnych adekwatne dla roślin młodych mogą być bezpośrednio przeniesione na ocenę roślin dojrzałych?
- 7) W podsumowaniu pracy brakuje jednoznacznego porównania wyników własnych analiz do tych otrzymanych przez innych autorów. Autor stosuje co prawda sformułowania porównawcze, np. „Niektórym autorom udało się uzyskać lepsze wartości metryk (...)”, lecz bez formalnego dowodu lub systematycznej próby zestawienia danych w kontekście specyfiki i warunków eksperymentu pomiarowego.
- 8) Wiele chorób spowodowanych przez patogeny, takie jak grzyby, bakterie i wirusy, ma bardzo podobne objawy i przebieg. Co więcej, wpływ na wygląd i ogólną kondycję roślin mają niedobory składników pokarmowych. Czy w ramach przyszłych badań rozważano analizę specyficzności wskazanych pasm hiperspektralnych, przy uwzględnieniu dużo większego zbioru obejmującego osobniczość gatunkową i chorobową?

5. Wnioski końcowe

Dokonując ogólnej oceny stwierdzam, że przedstawiona rozprawa doktorska mgra inż. Krzysztofa Smykały jest pracą wartościową, stojącą na wysokim poziomie naukowym. Zawartość dysertacji wskazuje, że została ona wykonana w obszarze dyscypliny naukowej *automatyka, elektronika, elektrotechnika i technologie kosmiczne*.

Praca zawiera szereg wymienionych wcześniej elementów oryginalnych. Przedstawione w recenzji uwagi krytyczne nie dotyczą zasadniczych walorów merytorycznych i tym samym nie pomniejszają wartości naukowych osiągnięć.

Cel pracy został osiągnięty, a postawiona na wstępie teza badawcza udowodniona. Przyjęta metodyka badań, zakres prac oraz uzyskane wyniki świadczą o pracowitości, zaangażowaniu, kreatywności i umiejętności samodzielnelnego prowadzenia badań naukowych przez Autora.

Biorąc pod uwagę powyższą ocenę dysertacji, wykazującą oryginalność rozwiązania postawionego problemu naukowego stwierdzam, że praca doktorska pt. „Zastosowanie metod uczenia maszynowego i pomiarów hiperspektralnych w diagnostyce uprawy *Solanum lycopersicum*” spełnia wymagania wynikające z przepisów ustawy Prawo o szkolnictwie wyższym i nauce z dnia 20 lipca 2018 r. (Dz.U z 2023 r. poz.742), w związku z czym **wnoszę o dopuszczenie mgra inż. Krzysztofa Smykały do publicznej obrony przedłożonej pracy.**

Gregorz Śrimalch