

Olsztyn, 2018-03-06

dr hab. inż. Piotr Zapotoczny, prof. UWM  
Katedra Inżynierii Systemów  
Wydział Nauk Technicznych  
Uniwersytet Warmińsko -Mazurski w Olsztynie  
e-mail: zap@uwm.edu.pl

**RECENZJA ROZPRAWY DOKTORSKIEJ**  
MGR ANNY SIEDLISKIEJ

**PT.: OKREŚLENIE WYBRANYCH CECH JAKOŚCIOWYCH OWOCÓW W OPARCIU O ANALIZĘ  
HIPERSPEKTRALNĄ ORAZ METODY SZTUCZNEJ INTELIGENCJI**

Niniejsza recenzja sporządzona została na zlecenie Rady Naukowej Instytutu Agrofizyki im. B. Dobrzańskiego Polskiej Akademii Nauk w Lublinie. Promotorem rozprawy doktorskiej jest dr hab. Piotr Baranowski, prof. IA PAN, kierownik Zakładu Metrologii i Modelowania Procesów Agrofizycznych, Instytutu Agrofizyki PAN, natomiast promotorem pomocniczym dr Monika Zubik.

Pani mgr Anna Siedliska przedstawiła do recenzji rozprawę doktorską pt.: *Określenie wybranych cech jakościowych owoców w oparciu o analizę hiperspektralną oraz metody sztucznej inteligencji*, składającą się z cyklu trzech wieloautorowych publikacji oraz uzupełnienia do badań. Publikacje zostały wydane w czasopismach naukowych z listy "A" Ministerstwa Nauki i Szkolnictwa Wyższego, punktowane w dacie ich opublikowania po 40, 40 i 25 punktów, co w przeliczeniu na procentowy udział Doktorantki daje 82 punkty ministerialne i 4,654 IF. W bazie Web of Science posiada 5 publikacji, indeks H -2 oraz 23 cytowania, jak na kandydatkę do uzyskania stopnia doktora dorobek jest zupełnie wystarczający i czyni zadość wymaganiu określonymu w art. 13 ust. 4 ustawy o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuk z dnia 14 marca 2003 roku. W załączonych dokumentach do rozprawy doktorskiej, zostały przedstawione oświadczenie współautorów publikacji o indywidualnym wkładzie każdego z nich w publikacje P.1, P.2 oraz P.3.

Praca ma nietypowy format będącym autorskim połączeniem dwóch form przedstawiania dysertacji tj.: cyklu publikacji oraz uzupełniona o klasyczną formę rozprawy (rozdział 3.4) jako uzupełnienia badań już opublikowanych. Ustawa o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuk z dnia 14 marca 2003 roku, art. 13 ust. 2 definiuje jaką formę może mieć rozprawa doktorska. Mając na uwadze powyższe, stwierdzam, że przedstawiona forma rozprawy doktorskiej spełnia zapisy wyżej cytowanej Ustawy.



Przedstawiona do oceny praca obejmuje bardzo aktualne zagadnienia z jakimi borykają się producenci owoców i warzyw, którzy to poszukują szybkich metod oceny surowca w czasie jego wegetacji oraz tuż po zbiorze. Także na etapie przetwórstwa, na liniach sortowniczych, potrzebujemy szybkich i niezawodnych technik, aby skutecznie sortować surowiec na klasy jakościowe.

Aktualne trendy w produkcji rolno-spożywczej ogniskują się nie na zwiększaniu plonów z jednostki powierzchni a na uzyskaniu surowca o wysokiej jakości. Wraz z ociepleniem klimatu coraz większym problemem jest porażenie surowca grzybami patogennymi. Jest to problem występujący już na polu w trakcie trwania wegetacji a także w czasie przechowywania. Także ważne dla praktyki rolniczej jest opracowanie taniej, szybkiej metody określania składu chemicznego surowca a w obrocie handlowym gotowego produktu. Dysponowanie takim technikami pozwoli na wykrycie zagrożeń mikrobiologicznych oraz zafałszowania składu chemicznego produktu jakiego mogą dopuścić się producenci, czy inni aktywni uczestnicy działający na rynku rolno-spożywczym.

Obowiązująca norma *General Marketing Standard*, zwana *Ogólną Normą Handlową* dla większości owoców i warzyw, określa minimalne wymagania dotyczące jakości i dojrzałości owoców, bez podziału na klasy. Natomiast szczegółowe normy zostały opracowane tylko dla produktów takich jak: jabłka, owoce cytrusowe, kiwi, sałata, endywia o liściach kędzierzawych i endywia o liściach szerokich, brzoskwinie i nektaryny, gruszki, truskawki, papryka słodka, winogrona stołowe, pomidory.

W ocenie jakości wykorzystujemy metody instrumentalne i organoleptyczne. Spośród metod instrumentalnych wyróżnić możemy metody niszczące jak i nieniszczące. Wśród metod nieniszczących mamy do dyspozycji komputerowe systemy wizyjne. Ostatnie lata to szczególnie rozwój metod wizyjnych opartych na rejestracji obrazu w zakresie bliskiej i dalekiej podczerwieni jak i ultrafioletu a w tym rejestracja obrazu w tzw. kanałach spektralnych zwanych potocznie multi-, super- lub hiperspektralne. Ten typ obrazowania niesie za sobą znacznie więcej informacji w stosunku do metod klasycznych z zakresu 400-700nm. Metody hiperspektralne umożliwiają między innymi identyfikację gatunków i odmian zbóż, stopnia zagrzybienia owoców oraz składu chemicznego, ma zastosowanie w ocenie jakości mięsa i wędlin. Jest to technika, która prawdopodobnie w najbliższych latach wyprze klasyczne obrazowanie i stanie się metodą referencyjną dla metod chemicznych lub mikrobiologicznych. Niestety na chwilę obecną jest metodą drogą w zastosowaniu oraz wymagającą dużej wiedzy na etapie jej wdrażania. Ilość i rodzaj uzyskanej informacji z pomiaru spektrofotometrycznego wymaga znajomości zagadnień związanych z analizą sygnału, analiz wielowymiarowych czy sztucznych sieci neuronowych.

### **Szczegółowa ocena dysertacji**

Przedstawiona praca składa się z następujących rozdziałów: (i) spisu symboli, (ii) listy publikacji stanowiących podstawę rozprawy doktorskiej, (iii) wprowadzenia, (iv) celu i hipotez badawczych, (v) omówienia wyników cyklu publikacji, (vi) badań uzupełniających, (vii) wniosków, oraz załączonych publikacji i bibliografii. Całość liczy 98 ponumerowanych stron. We wprowadzeniu autorka na 29 stronach przedstawiła przegląd literatury na temat wskaźników oceny jakości owoców i warzyw oraz metod ich pomiaru. Omówiła szczegółowo zasady działania systemów hiperspektralnych oraz ich zastosowania w ocenie jakości owoców.





W tabeli 1.1 przedstawiła syntetycznie dane dotyczące wyników badań z oceny jakości owoców technikami hiperspektralnymi. Jest to o tyle cenne, że przegląd obejmuje nie tylko informacje o zakresie prowadzonych badaniach a także jakie wykorzystano metody statystyczne w analizie danych. Ostatnie dwa rozdziały przeglądu literatury autorka przeznaczyła na omówienie metody analizy obrazu oraz selekcji danych hiperspektralnych oraz metodach klasyfikacji danych hiperspektralnych. Omówiła w skrócie metody wstępnego przetwarzania jak: transformacje SNV (ang. Standard Normal Variate), algorytm MSC (ang. Multiplicative Scatter Correction), algorytm Savitzky-Golay. Oczywiście nie są to jedyne metody, ale rodzaj pracy nie wymagał od autorki szczegółowych opisów wszystkich technik wstępnego przetwarzania sygnału np. transformacji „offset correction”.

W kolejnym akapicie omówiła zagadnienie segmentacji obrazu, czyli podziału obrazu na piksele tła oraz interesującego nas obiektu. W mojej ocenie jest to rozdział zbyt krótki, autorka mogłaby by te zagadnienie omówić znacznie szerzej. Segmentacja obrazu nie jest sprawą trywialną i może być kluczowa w całym algorytmie przetwarzania obrazu i jego analizy.

Ostatnim zagadnieniem poruszonym przez Autorkę było omówienie metod klasyfikacji nadzorowanej i nie nadzorowanej przeznaczonych do analizy danych hiperspektralnych. Omówiła metodę PCA, w skrócie wspomniała o metodach nadzorowanych LDA, PLS-DA, kNN. Omówiła także zagadnienie związane z wielowymiarową analizą regresji (metoda MLR -wielokrotną regresję liniową, PCR - regresją składowych głównych czy PLSR regresję najmniejszych kwadratów). Klika zdań zostało poświęcone sztucznym sieciom neuronom ANN, SVM. Według mojej opinii rozdział ten jest mało wyczerpujący, tym bardziej, że jednym z członów tytułu prezentowanej dysertacji jest wykorzystanie metod sztucznej inteligencji.

W dalszej części pracy autorka przedstawiła cel rozprawy doktorskiej i hipotezy badawcze, formułując cztery hipotezy oraz pięć celów rozprawy.

Hipotezy badawcze są zdaniami, które wymagają sprawdzenia, w celu uznania ich za fałszywe lub prawdziwe. Warunkiem uznania zdania za hipotezę jest sprawdzenie czy jest:

1. nowa, czyli zawierająca jakieś niezbadane dotychczas kwestie lub aspekty,
2. ogólna, tzn. musi brać pod uwagę wszelkie fakty, procesy i zjawiska związane z badaną kwestią,
3. pojęciowo jasna, czyli wyrażona dokładnie i jednoznacznie,
4. niesprzeczna wewnętrznie,
5. sprawdzalna,
6. nie oczywistością lub banałem.

Mając powyższe na uwadze, uważam, że postawione hipotezy badawcze są zbyt oczywiste. Czytając z uwagą przegląd literatury (patrz. tabela 1.1) można byłoby sformułowane hipotezy sfalsyfikować bez prowadzenia badań.

Przykład:

Hipoteza nr 1: *Zróznicowanie właściwości wewnętrznych i przypowierzchniowych tkanki owoców, wywołane np. uszkodzeniami mechanicznymi, intruzjami wewnętrznymi lub porażeniem przez patogeny grzybowe, może być skutecznie wykrywane i klasyfikowane przy*

wykorzystaniu systemów hiperspektralnych pracujących w zakresach VNIR/SWIR oraz w trybie odbiciowym lub transmisyjnym.

Odpowiedź, bez prowadzenia badań, brzmi tak - patrz tabela 1.1 (Dong i Guo, 2015, Zhang i in., 2015a, Fan i in., 2017, Lee i in., 2014).

Podobnie można postąpić z hipotezą nr 2 oraz 3. Jedynie hipoteza badawcza nr 4 nie ma oczywistej odpowiedzi w prezentowanym przeglądzie literatury, ale jest niefortunnie sformułowana. Czy klasyfikujemy różnice odmianowe, czy odmiany?

Sugerowałbym w przyszłym przygotowaniu materiału do publikacji przereformowanie hipotez badawczych.

Natomiast nie mam zastrzeżeń, co do poprawności założonych celów rozprawy doktorskiej, zostały sformułowane precyzyjnie i jasno, definiują zamierzenia badawcze jakie postawiła sobie Doktorantka.

**W rozdziale 3** Pani mgr Anna Siedliska omówiła wyniki badań zaprezentowanych w publikacjach wchodzących do cyklu oraz badaniach uzupełniających. W badaniach jako materiał badawczy użyła owoców jabłek, wiśni oraz truskawek. Techniki obrazowania hiperspektralnego wykorzystwała do identyfikacji uszkodzeń mechanicznych, obecność pestek, zawartości ekstraktu oraz wczesnej identyfikacja chorób wywołanych przez patogeny grzybowe. Na uwagę zasługuje podejście do analizy statystycznej wyników. Doktorantka wykorzystwała szeroki zakres metod klasyfikacji nadzorowanej takiej jak:

- modele statystyczne: Naive Bayes (NB)
- modele liniowe: Logistic (LOG), Simple Logistic (SLOG), LibLINEAR (LINE)
- drzewa decyzyjne: Random Forest (RF)
- modele sieci neuronowych: Multilayer Perceptron (BNN)
- modele funkcji jądrowych: metoda wektorów wspierających (SVM), algorytm SMO
- modele oparte na interakcjach: IB1
- model logiki rozmytej: FURIA

**W publikacji P.1** przedstawiona została możliwość zastosowania metody obrazowania hiperspektralnego w zakresach VNIR/SWIR do stworzenia modeli klasyfikacji służących do wykrywania uszkodzeń mechanicznych (obić) w jabłkach. Model klasyfikacyjny skuteczności klasyfikacji uszkodzeń został zbudowany na bazie wyselekcjonowanych 33 długości fali. Niezależnie od zastosowanej metody klasyfikacji ilość poprawnie sklasyfikowanych przypadków zawierała się w przedziale od 73,3 do 90%, przy czym najlepszą zdolność predykcji uzyskano dla modeli SVM, SLOG oraz SMO. W tej samej pracy zaprezentowano także wyniki, klasyfikacji odmianowej jabłek, uzyskując poprawność klasyfikacji na poziomie 93%.

**W publikacji P.2** potwierdzono użyteczność metody obrazowania hiperspektralnego w trybie światła przechodzącego do identyfikacji niejednorodności (obecność pestek oraz ich fragmentów) wewnątrz tkanek owoców wiśni. Autorzy wykazali, że modele klasyfikacyjne, bazujące na wartościach współczynnika odbicia dla 19 lub 20 wybranych długości fal, uzyskane za pomocą algorytmu CFS, są dobrym narzędziem do identyfikacji pestek oraz ich fragmentów w badanych owocach.



**W publikacji P.3** autorka dalej kontynuowała prace związane z opracowaniem metody identyfikacji pestek w owocach wiśni oraz predykcji zawartości ekstraktu w owocach. Skonstruowane modele klasyfikacyjne nadzorowane uwzględniające 27 zmiennych pozwalały na poprawne sklasyfikowanie, co najmniej 85% przypadków. Wyjątek stanowił model NB, dla którego uzyskano skuteczność na poziomie 57%. Szacowanie zawartości ekstraktu zrealizowano budując model regresyjny na bazie pięciu głównych składowych co pozwoliło na otrzymanie modelu, w którym współczynnik determinacji  $R^2$  wynosił 0,636, natomiast RMSEP (pierwiastek błędu średniokwadratowego) był równy 0,93.

**Rozdział 3.4** przedstawionej do oceny pracy zawierał badania uzupełniające pt.: *Wykrywanie porażień grzybowych truskawek z wykorzystaniem obrazowania hiperspektralnego w zakresach VNIR/SWIR.*

Materiałem badawczym były odmiany 'Senga Sengana' oraz 'Honeoye', które zostały inokulowane zarodnikami 2 gatunków grzybów (*Botrytis cinerea* oraz *Colletotrichum acutatum*). Badania hiperspektralne zostały zrealizowane na stanowisku składającym się z kilku kamer, pozwalających na zarejestrowanie obrazu w zakresie od 360 do 2569nm. Analiza statystyczna wyników objęła segmentację obrazu, redukcję zakłóceń wynikającą z nierównomiernego oświetlenia próbek w czasie pomiaru, analizę klasyfikacyjną, analizę fizykochemiczną oraz analizę regresji. W podsumowaniu autorka potwierdziła przydatność zastosowanych metod do identyfikacji stopnia porażenia patogenami wybranych odmian truskawek, jak i zaproponowała równania regresji do predykcji zawartości ekstraktu, antocyjanów lub związków fenolowych na podstawie pomiar kilku długości fal. Układ doświadczalny oraz przeprowadzona analiza statystyczna wyników zostały dobrane poprawnie, a uzyskane wyniki jasno potwierdziły poprawność założeń badawczych.

Opisane osiągnięcie naukowe ma charakter wielowątkowy. Do najważniejszych osiągnięć naukowych Doktorantki zaliczam:

- Opracowanie modeli statystycznych do identyfikacji porażenia grzybami *Botrytis cinerea* i *Colletotrichum. acutatum* owoców truskawek,
- Opracowanie modeli klasyfikacji uszkodzeń mechanicznych,
- Opracowanie modeli klasyfikacji wybranych odmian jabłek,
- Zaproponowanie metody identyfikacji pestek i ich fragmentów w owocach wiśni,
- Wyjaśnienie powiązań pomiędzy zawartością ekstraktu oraz antocyjanów z wybranymi kanałami spektralnymi w pomiarach hiperspektralnymi. Równania regresji zostały opracowane dla owoców truskawek.

#### **Uwagi merytoryczne:**

1. Nie jest jasna metodyka, co do ilości sfotografowanych obiektów w każdej grupie doświadczalnej. Autorka podaje, że wybrano 2700 szt. owoców należących do dwóch

odmian. Następnie podzielono je na trzy grupy; kontrolną, inokulowane grzybem BC oraz grzybem CA. Brak informacji czy podział nastąpił w równej proporcji. Domyślam się, że podział jest następujący:  $2700/2=1350$  jednej odmiany/ $3=450$  szt. dla jednej grupy doświadczalnej?

Następnie w podrozdziale 3.4.2.3 podane jest, iż pojedynczy obraz zawierał 10 owoców, czyli można domyślać się, że wykonano 45 zdjęć każdej grupie doświadczalnej. Niestety metodyka tego dokładnie nie opisuje i warto byłoby to sprecyzować.

2. Na stronie 40 (podrozdział 3.4.3 Wyniki), Doktorantka użyła cyt.: *Na rys. 3.1 przedstawiono przykładowe obrazy hiperspektralne badanych owoców truskawek uzyskane dla kompozycji trzech barw (R: 638 nm, G: 549 nm, B: 458 nm).*  
Należało użyć nie barwy, tylko kanały. Barwa powstaje w wyniku "mieszania" składowych barwnych, w przypadku modelu RGB, jest to "mieszanie" tych trzech kanałów.
3. Na Rys.3.2 nie wyjaśniono co oznacza legenda na wykresie, podano wartości od 0 do 4. Mogę domyślać się, że są to kolejne krzywe spektralne dla n-powtórzenia lub uśrednione krzywe spektralne dla poszczególnych dni, ale nie jestem pewien.
4. W podrozdziale 3.4.3.2 Autorka pisze cyt.: *Analiza klasyfikacyjna danych spektralnych została przeprowadzona w dwóch etapach. Pierwszy dotyczył podziału owoców na trzy grupy: owoce zdrowe, inokulowane grzybem BC oraz grzybem CA*, dalej podaje liczebności poszczególnych grup. I tu nasuwa się pytanie: czy faktycznie podział na grupy był pierwszym etapem analizy klasyfikacyjnej?
5. W podrozdziale 3.4.3.2 autorka pisze cyt.: *Przedstawione w Tab. 3.1 wyniki potwierdzają wysoką dokładność klasyfikacji próbek testowych, która waha się w zakresach 70–88% dla odmiany 'Senga Sengana', 78–90% dla odmiany 'Honeoye' oraz 63–87% dla obu odmian.* Co to jest próbka testowa?, chyba chodziło o zbiór testowy?
6. W podrozdziale 3.4.3.2 doktorantka użyła następującego stwierdzenia: cyt.: *najlepszą metodą klasyfikacji okazał się model BNN*. Czy faktycznie był to model BNN?, czy raczej model klasyfikacyjny został zbudowany lub wygenerowany za pomocą sieci neuronowych z zastosowaniem algorytmu BNN?  
I dalej w tym samym podrozdziale, cyt.: *Jak pokazuje zestawienie, najlepszą metodą klasyfikacji okazał się model BNN, który przypisuje badane obiekty z dokładnością wyższą niż 87%.*  
Narzuca się pytanie, do czego przypisuje badane obiekty, czy do grupy pozytywnej czy negatywnej. Może warto byłoby użyć stwierdzenia, że poprawność klasyfikacji wyniosła ponad 87%, jest to niuans, ale precyzuje co jest kryterium klasyfikacji.
7. Proponuję ujednoczyć przedstawianie wyników tj. podawany % dokładności klasyfikatora. Raz Autorka podaje wynik z dokładnością bez przecinka raz z dokładnością do dziesiątej. Wydaje się, że wartość po przecinku nie jest liczbą znaczącą, ale to pozostawiam do decyzji Doktorantki.



8. Tab. 3.1. Brak jest informacji z jakiego dnia po inokulacji zostały przedstawione wyniki i dla jakiego zbioru.

9. Nie zrozumiałam jest podpis pod Rys. 3.3; cyt.: *Porównanie danych sklasyfikowanych z danymi referencyjnymi (zmierzonymi) dla klasyfikatora BNN i modelu służącego do klasyfikacji owoców zdrowych oraz owoców porażonych grzybem BC oraz CA.*

Co to są dane referencyjne a dane sklasyfikowane? oraz czy klasyfikator BNN jest tym samym co model służący do klasyfikacji owoców? Także mam znów wątpliwości, co do liczebności zbiorów, jeżeli przyjąć, że jest to zbiór testowy, który stanowi 25% całego zbioru to cała próba powinna liczyć  $n=1280$  dla grupy BC oraz  $n=1280$  co daje łącznie  $n=2560$  co jest niezgodne z liczebnością podaną w metodyce.

10. Rys.3.4 Nie jest to zarzut, ponieważ świadomy jestem, że są to wyniki otrzymane metodami statystycznymi, przy założonym poziomie istotności wnioskowania statystycznego, ale proszę spróbować odpowiedzieć na pytanie: jak należy interpretować wynik z macierzy literka "d", w 3 dniu po inokulacji jeden przypadek został niepoprawnie sklasyfikowany i przypisany do dnia 2, natomiast po 4 dniu, siedem przypadków zostało przypisanych do dnia 2, czy nastąpiło cofnięcie się porażenia grzybami?

Należy jednak przyznać, iż uzyskane wyniki są bardzo obiecujące, i pozwalają na kontynuowanie badań w celu dopracowania metody identyfikacji porażenia patogenami owoców miękkich.

11. Rys. 3.5 Brak jest w legendzie wyjaśnienia, co oznaczają pionowe słupki, czy jest to błąd standardowy, błąd standardowy średniej, czy inna statystyka.

Brakuje informacji, jakie było  $p=$  dla składowych w modelu regresyjnym.

12. Strona 16. We wzorze na współczynnik reflektanci występuje błąd w oznaczeniu,  $I_m$  w oznaczeniu, powinno być  $I_{im}$ .

Uwaga ogólna do rozdziału 3.4. Brak szczegółowych informacji na temat metod analizy statystycznej. Nie podano dla sieci neuronowych, które były najskuteczniejsze, ilości neuronów wejściowych, ilości warstw ukrytych, podziału na zbiór uczący, walidacyjny i testowy. W przypadku metody RF (Random Forest) oraz NB (Naive Bayes) nie podano konkretnej metody, program WEKA, w tych grupach zawiera po kilkanaście konkretnych metod. Ponadto nie opisano czy zastosowano walidację krzyżową czy % podział zbioru danych itp. W tytule pracy doktorskiej położony jest nacisk na metody sztucznej inteligencji, z tego względu powinno się na te zagadnienie poświęcić więcej miejsca wraz z bardziej szczegółowym opisem metod.

## Podsumowanie

Recenzowana rozprawa doktorska mgr Anny Siedliskiej pt.: **"Określenie wybranych cech jakościowych owoców w oparciu o analizę hiperspektralną oraz metody sztucznej**

**inteligencji"** stanowi samodzielne rozwiązanie problemu badawczego z obszaru nauk rolniczych w dyscyplinie agronomii. Za szczególne istotne należy uznać podjęcie badań w zakresie opracowania metody identyfikacji zagrzybienia owoców nietrwałych na przykładzie truskawek. Autorka wykazała się wysoką znajomością literatury przedmiotu oraz dobrym przygotowaniem do planowania i realizacji eksperymentów badawczych. W mojej ocenie cele badań zostały zrealizowane w pełni, autorka wniosła nowe elementy poznawcze i rozwiązała kilka problemów badawczych będących istotnymi dla dziedziny nauk rolniczych. W świetle powyższego stwierdzam, że przedłożona do recenzji rozprawa doktorska spełnia ustawowe kryteria stawiane pracom doktorskim zgodnie z Ustawą o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. 65 poz. 595 z dnia 14 marca 2003 roku) i na tej podstawie stawiam wniosek o dopuszczenie mgr Anny Siedliskiej do dalszej części przewodu doktorskiego.

A handwritten signature in blue ink, appearing to read 'Piotr Zapotoczny', written over a horizontal dashed line.

dr hab. inż. Piotr Zapotoczny, prof. UMW