

Dokumentacja Projektu

Dostrajanie klasyfikatora FUZZY LOGIC

z użyciem DGO

Adam Łaba

Julia Bała

Kacper Goraj

13 czerwca 2022

1. Wstęp

Projekt polega na utworzeniu układu Fuzzy Logic oraz następnie na dostrojeniu układu przy wykorzystaniu wybranego algorytmu optymalizacyjnego. Wybrany przez nas algorytm, pierwotnie wykorzystany do optymalnego rozłożenia kondensatorów w sieci rozdzielczej, którego autorami są Mohammad Dehghaniego, Zeinab Montazeri, Om Parkash Malik to Dice Game Optimization (DGO), szczegółowo opisany w:

<https://dergipark.org.tr/en/pub/gujs/issue/48219/484643>

2. Algorytm DGO

DGO inspirowany jest starą turecką grą w kości, której zasady są przedstawione poniżej.

Pole do gry składa się z pól. Każde pole ma przypisany do siebie konkretny wynik. Gracze rozmieszczeni są na polach (więcej niż jeden gracz może znajdować się na danym polu). Więc, każdy gracz ma dwa atrybuty: wynik rzutu kością i pole, na którym się znajduje. Gracze nie mają pojęcia o wynikach rzutu kością innych graczy.

Wynik rzutu kością danego gracza, to liczba jego '*przewodników*'. Każdy gracz w każdej turze może mieć od 1 do 6 *przewodników*. *Przewodnicy* są wybierani losowo spośród pozostałych graczy. W każdej turze, gracz będzie porównywał wynik swojego pola z wynikami pól swoich *przewodników*, i na tej podstawie wybierane jest pole, na jakim gracz znajdzie się w następnej turze.

W oparciu o powyższe powstał algorytm DGO. Przedstawia go poniższy pseudokod:

- 1 Inicjalizacja pozycji każdego z graczy
- 2 Pętla iteracyjna:
 - 3 Obliczenie wyniku każdego gracza
 - 4 Każdy gracz rzuca kością
 - 5 Losowanie przewodników dla każdego gracza
 - 6 Aktualizacja pozycji graczy
- 7 Koniec pętli
- 8 Koniec

Inicjalizacja pozycji każdego z graczy

W zależności od problemu, do jakiego wykorzystuje się optymalizację, początkowe pozycje wszystkich graczy mogą być losowane lub predefiniowane.

Pętla iteracyjna

Po wejściu do pętli, algorytm będzie powtarzać kolejne kroki aż do spełnienia warunku stopu.

Obliczenie wyniku każdego gracza

W algorytmie, wynik S dla każdego gracza nie jest bezpośrednio brany z wartości pola, na jakim się znajduje, tylko jest wyliczany przy wykorzystaniu wzoru:

$$S_i = \frac{fit_i - fit(worst)}{\sum_{j=1}^N fit_j - fit(worst)} \quad (1)$$

gdzie:

S_i - wynik dla gracza i

fit_i - wartość funkcji dopasowania dla gracza i

$fit(worst)$ - wartość funkcji dopasowania dla najgorszego gracza w stawce

Każdy gracz rzuca kością

Dla każdego z graczy, losowana jest liczba oczek (symulacja rzutu kością), która określa, ilu przewodników będzie mieć dany gracz.

Losowanie przewodników dla każdego gracza

Dla każdego gracza, zostaje losowo wybranych D przewodników spośród wszystkich pozostałych graczy, gdzie D to wartość rzutu kością danego gracza w danej iteracji.

Aktualizacja pozycji graczy

Przed końcem każdej iteracji, dla każdego z graczy obliczana jest pozycja w następnym kroku w oparciu o wzór:

$$X^i = X_0^i + \sum_{k=1}^{D_i} (r_k (X^i - X^{Guide_k}) \text{sign}(S_i - S_{Guide_k})) \quad (2)$$

gdzie:

X^i - Pozycja gracza i

D_i - wynik rzutu kością gracza i

r_k - liczba losowa pochodząca z rozkładu normalnego w zakresie $[0, 1]$

X^{Guide_k} - pozycja k przewodnika gracza i

S_i - wynik gracza i

S_{Guide_k} - wynik k przewodnika gracza i

Koniec pętli

Praca opisująca DGO nie definiuje warunku stopu, jest on zależny od implementacji.

W naszym rozwiązaniu mamy dwa warunki stopu - maksymalna liczba iteracji jest ograniczona, a także DGO zostanie przerwane jeśli wynik nie poprawił się w ostatnich n iteracjach.

3. Wykonanie zadania

Całość zadania została wykonana z wykorzystaniem języka MATLAB. Tworzone przez nas układy FL miały za zadanie dokonać klasyfikacji danych pochodzących ze zbiorów IRIS, WINE oraz SEEDS.

Pierwszym krokiem było wczytanie danych do programu.

Zbiór IRIS - 4 cechy, na podstawie których klasyfikujemy element do jednej z 3 klas.

Zbiór WINE - 13 cech, na podstawie których klasyfikujemy element do jednej z 3 klas.

Zbiór SEEDS - 7 cech, na podstawie których klasyfikujemy element do jednej z 3 klas.

W przypadku zbioru danych SEEDS, usunięte zostały wszystkie wpisy, w których co najmniej jedna z kolumn miała wartość NaN.

Następnie, podzieliliśmy losowo dane na zbiory uczące (70%) i testowe (30%).

Inicjalizujemy układ FIS za pomocą danych uczących, a następnie sprawdzamy, jaki rezultat osiąga dla danych testowych. Do utworzenia FIS użyliśmy algorytmu Subtractive Clustering.

Następnie pobieramy parametry układu FIS i optymalizujemy je przy wykorzystaniu DGO.

Na bazie otrzymanych wyników dostrajamy układ.

Ostatnim krokiem jest ponowne sprawdzenie, jak dostrojony FIS klasyfikuje dane testowe.

Przeprowadzony został następnie sprawdzian krzyżowy metodą CV-5. Podzieliliśmy nasz zbiór uczący na 5 podzbiorów, w oparciu o które została wykonana walidacja.

4. Uzyskane wyniki

Rezultaty przed walidacją:

4.1 IRIS

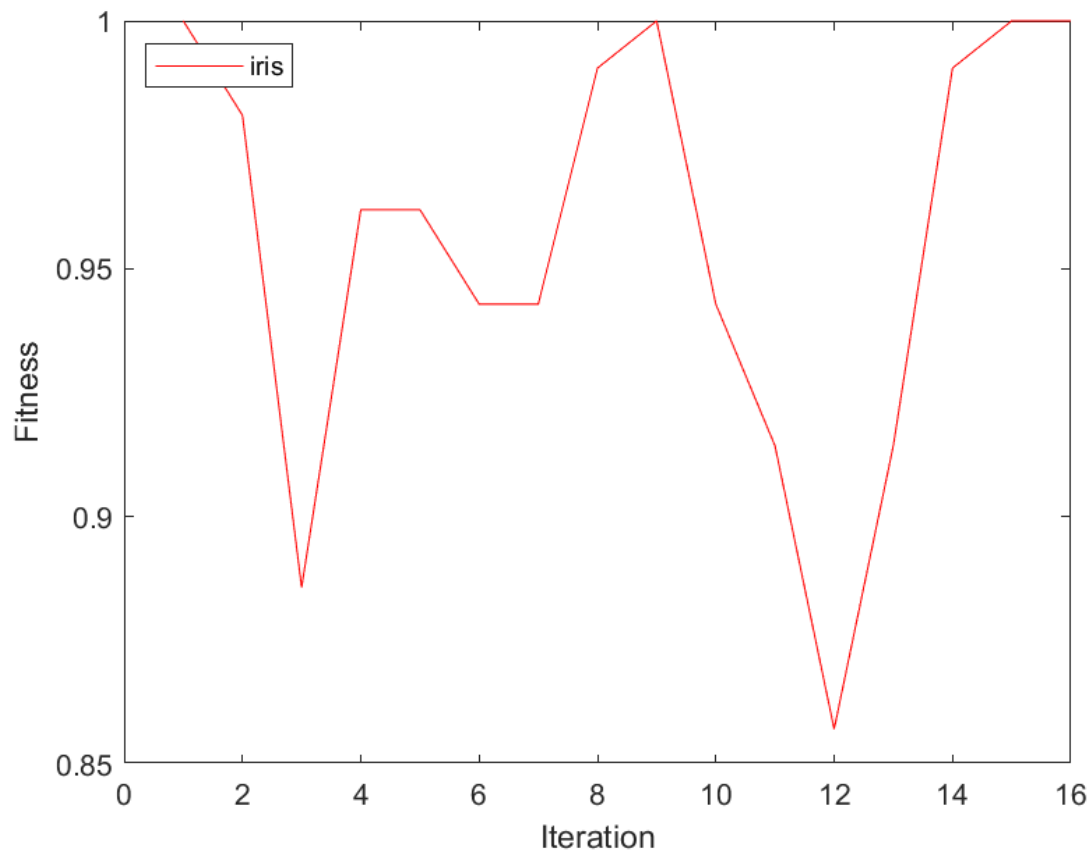
Dla tego zestawu otrzymaliśmy najlepsze wyniki.

Poziom dopasowania domyślnego FIS dla zestawu trenującego: 0,9905

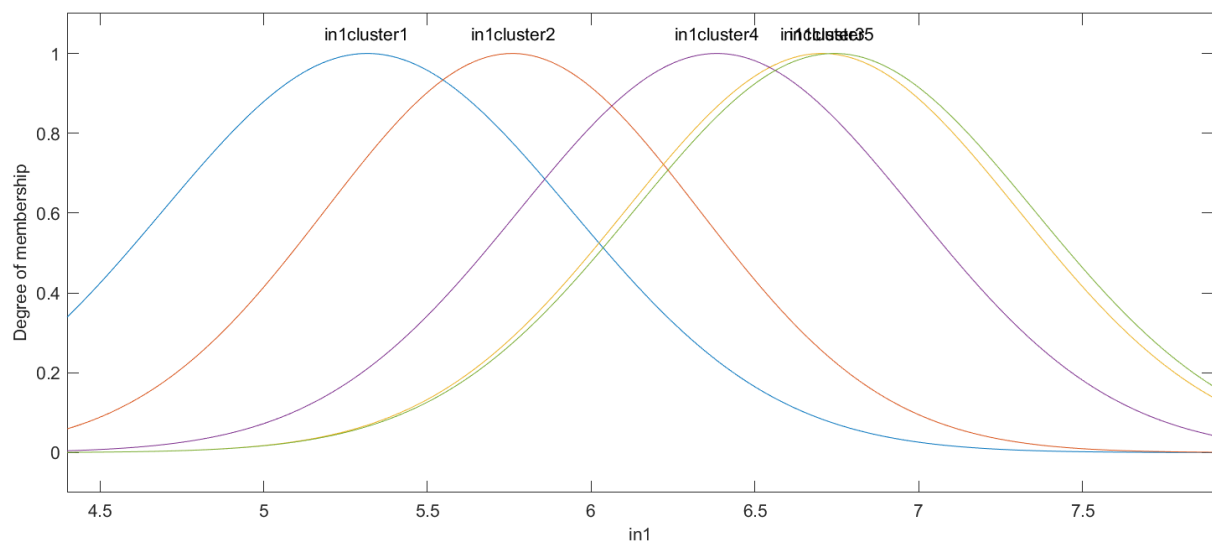
Poziom dopasowania domyślnego FIS dla zestawu testowego: 0,9556

Poziom dopasowania dostrojonego FIS dla zestawu trenującego: 1

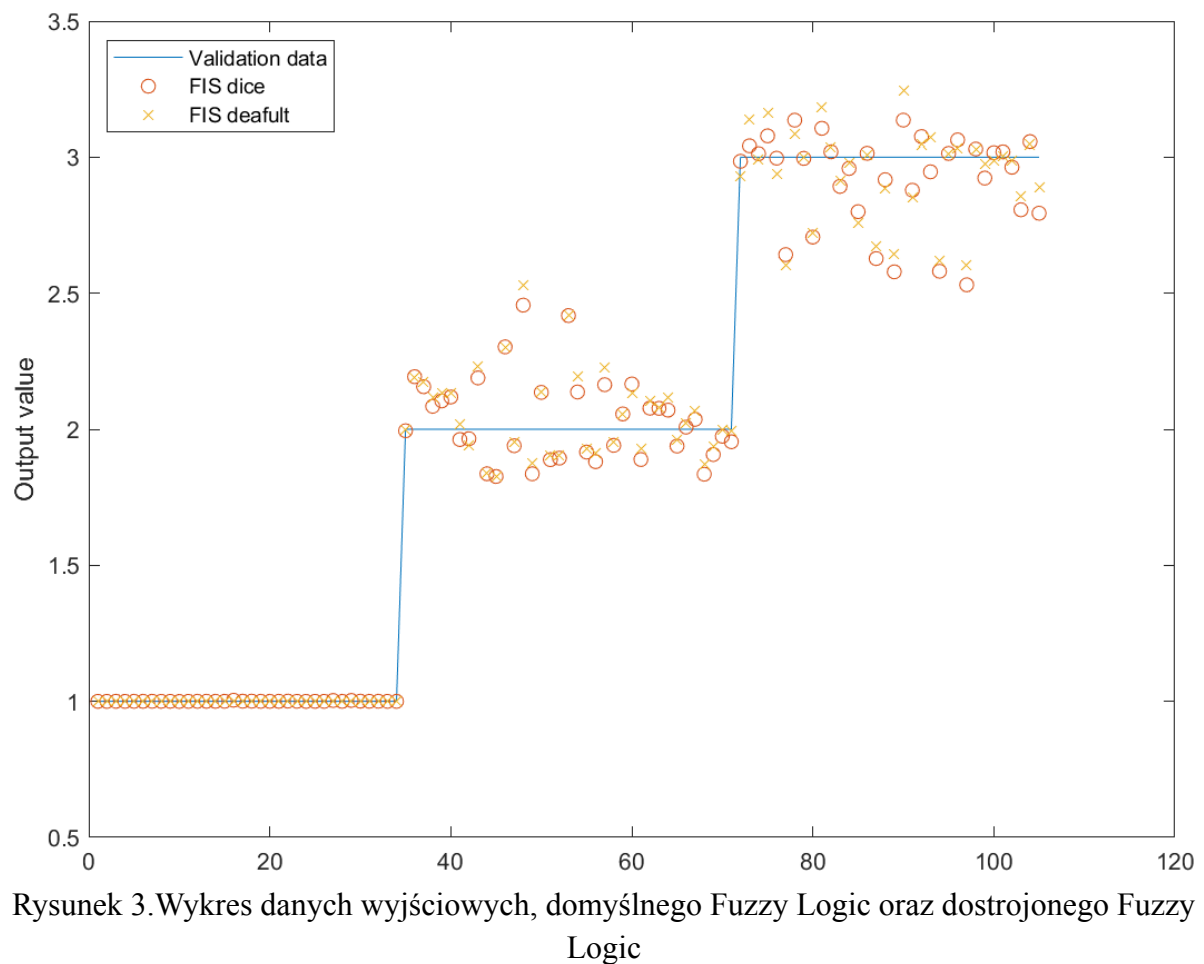
Poziom dopasowania dostrojonego FIS dla zestawu testowego: 0,9556



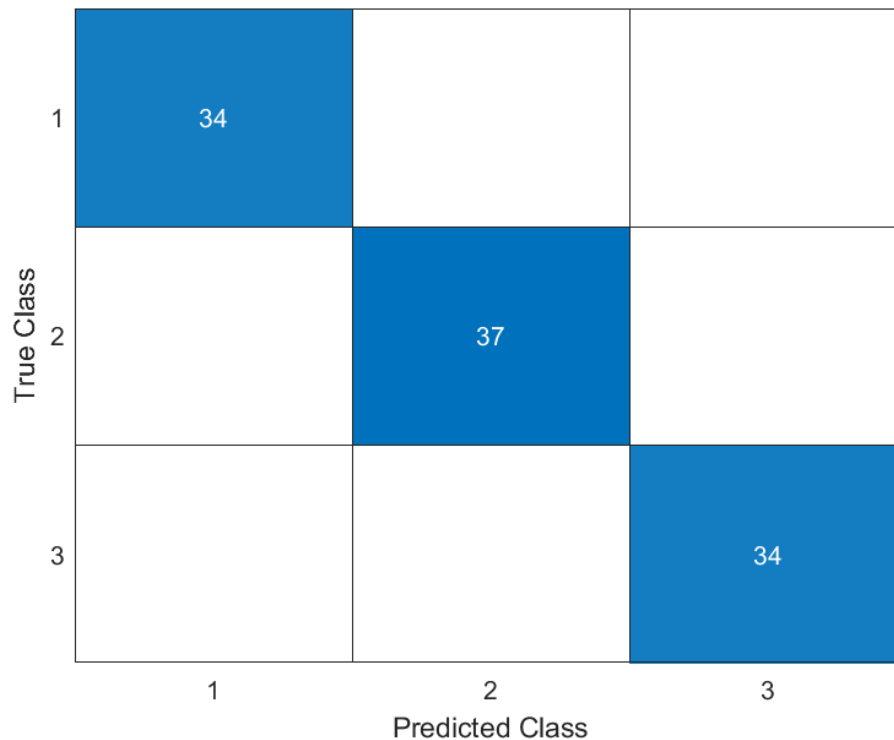
Rysunek 1 . Wykres najlepszego przystosowania w każdej iteracji.



Rysunek 2. Przykładowy wykres przynależności wejścia do rozmytego system wnioskowania



Rysunek 3. Wykres danych wyjściowych, domyślnego Fuzzy Logic oraz dostrojonego Fuzzy Logic



Rysunek 4. Macierz pomyłek dostrojonego układu FL dla zbioru Iris

Macierz pomyłek pozwala nam na wyznaczenie takich wartości jak:

- czułość (recall) i swoistość (specificity) -wartości opisujące zdolność testu do wykrycia badanej cechy (czułość) lub wykrycia jej braku (swoistość).
- dokładność (accuracy)- stopień zgodności wartości rzeczywistej ze średnią arytmetyczną wyników uzyskanych dla oznaczanej wielkości.
- precyzja (precision)- stopień zgodności między wynikami uzyskanymi w określonych warunkach z wielokrotnych pomiarów tej samej wielkości.

W przypadku macierzy pomyłek dla wielu klas, wyliczyć możemy:

- dokładność

$$accuracy = \frac{\sum_{i=1}^k e[i,i]}{\sum_{i=1}^k \sum_{j=1}^k e[i,j]}$$

- precyzję (dla poszczególnych klas)

$$precision(c_i) = \frac{e[i,i]}{\sum_{j=1}^k e[j,i]}$$

- czułość (dla poszczególnych klas)

$$recall(c_i) = \frac{e[i,i]}{\sum_{i=1}^k e[i,i]}$$

W przypadku zbioru Iris powyższe wartości wynoszą kolejno:

$$accuracy = \frac{34+37+34}{34+37+34} = 1$$

$$precision(c_1) = \frac{34}{34} = 1$$

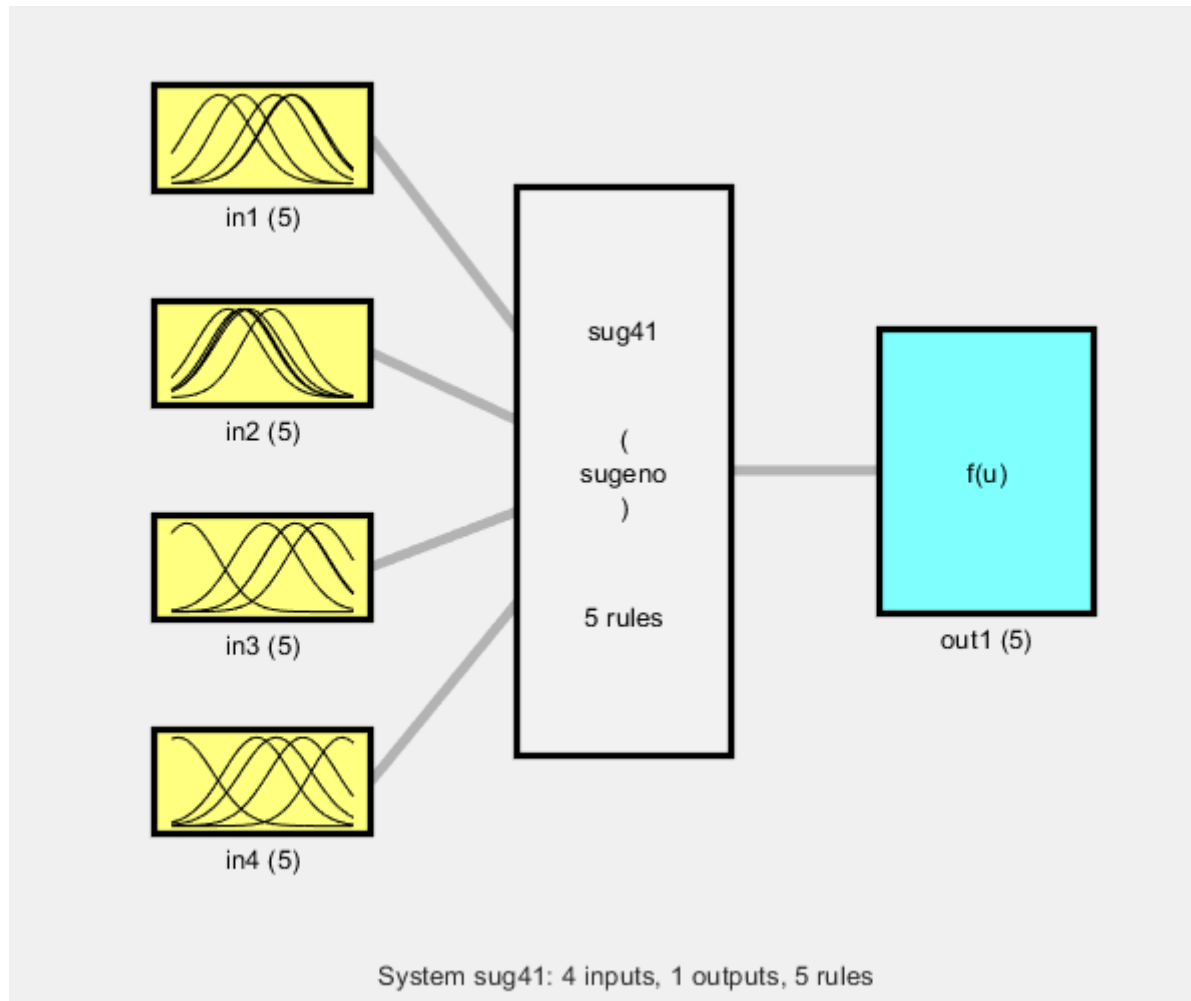
$$precision(c_2) = \frac{37}{37} = 1$$

$$precision(c_3) = \frac{34}{34} = 1$$

$$recall(c_1) = \frac{34}{34} = 1$$

$$recall(c_2) = \frac{37}{37} = 1$$

$$recall(c_3) = \frac{34}{34} = 1$$

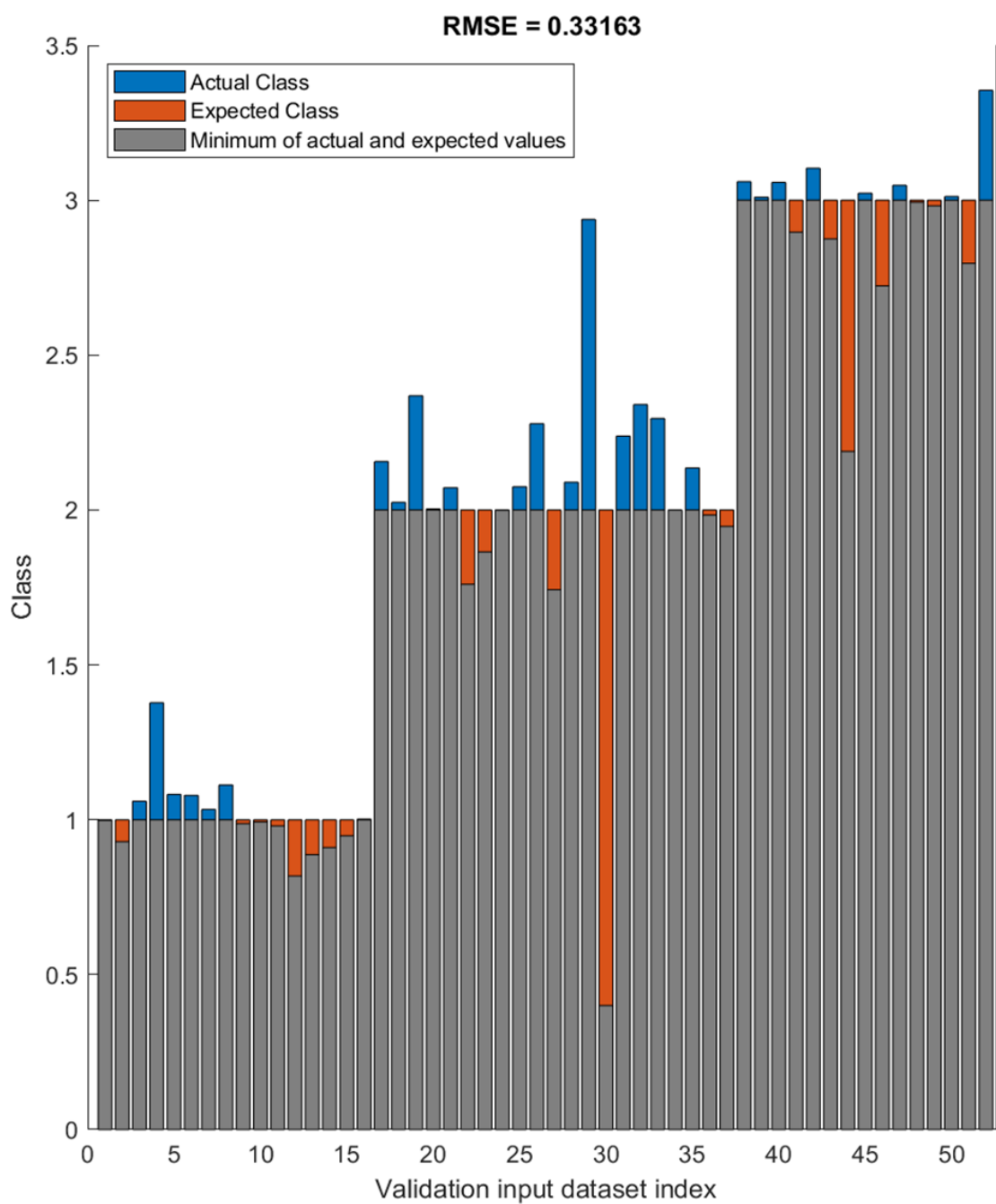


Rysunek 5. Struktura dostrojonego FIS.

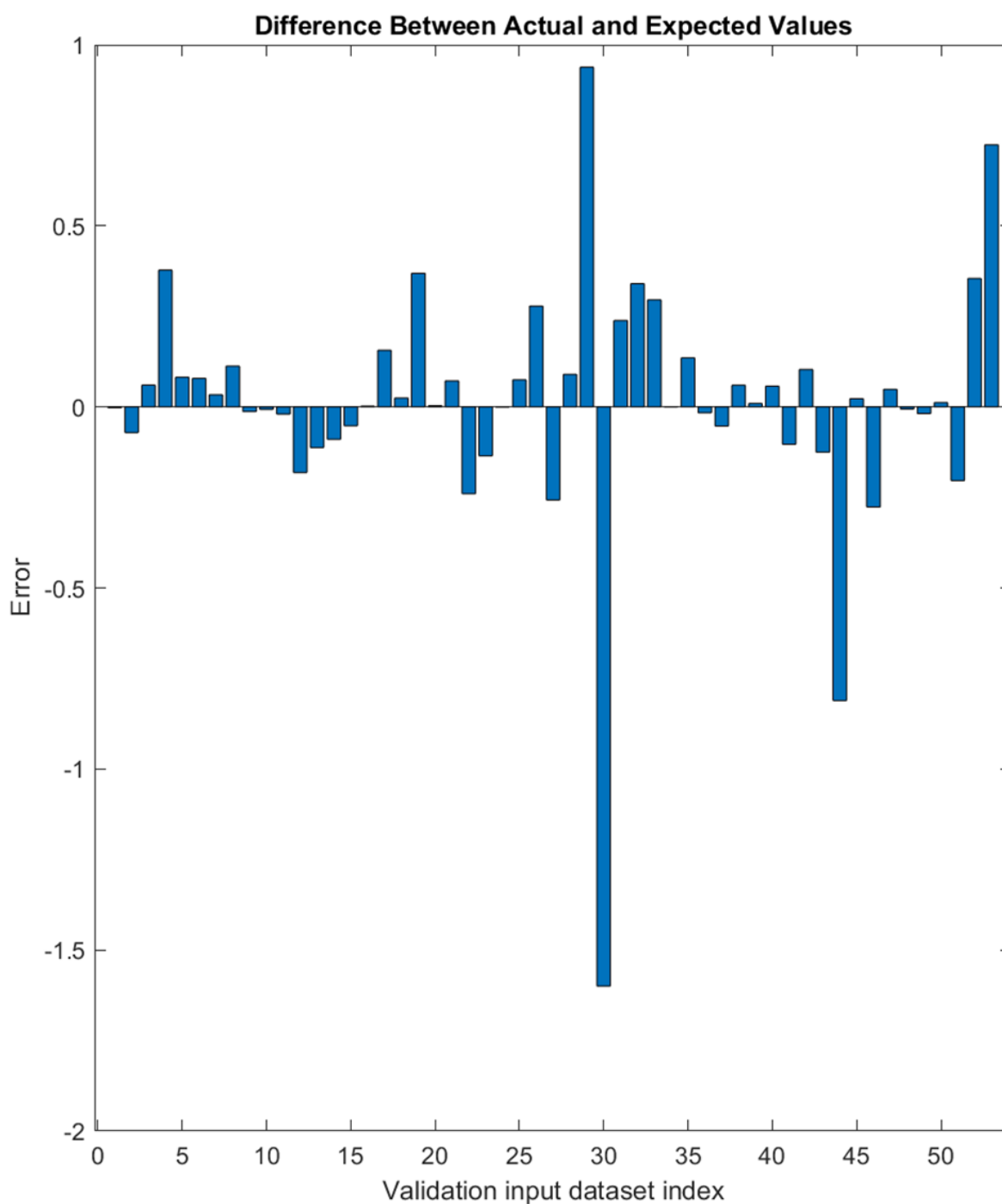
Średniokwadratowy błąd naszego dostrojonego układu Fuzzy Logic dla zestawu IRIS wyszedł niewielki. Wynik na zbiorze trenującym:

```
>> wykresy
```

```
Training RMSE = 0.141
```



Rysunek 6. Wykres otrzymanego i oczekiwanego wyniku z RMSE na zbiorze testującym



Rysunek 7. Wykres różnicy między otrzymanym a oczekiwanym wynikiem.

4.2 WINE

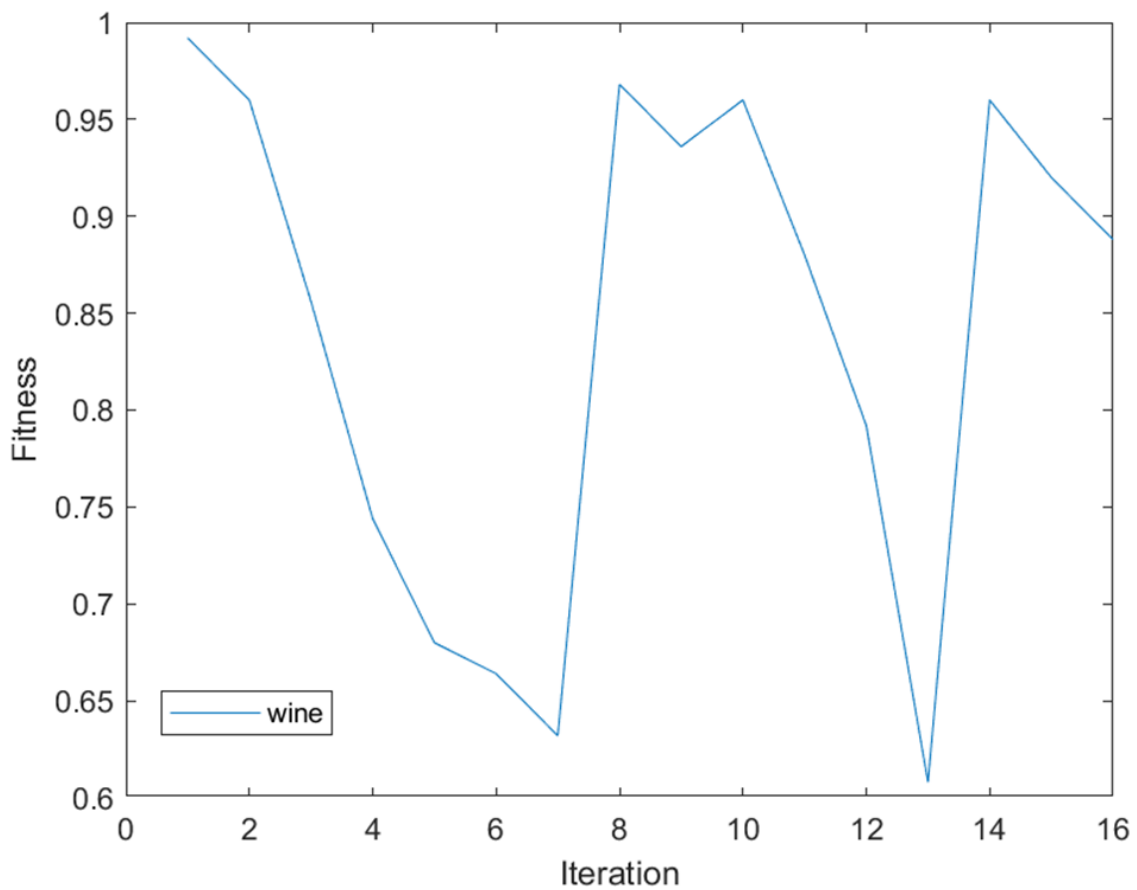
Dla tego zestawu otrzymaliśmy wyniki generowały się bardzo długo (24h).

Poziom dopasowania domyślnego FIS dla zestawu trenującego: 1

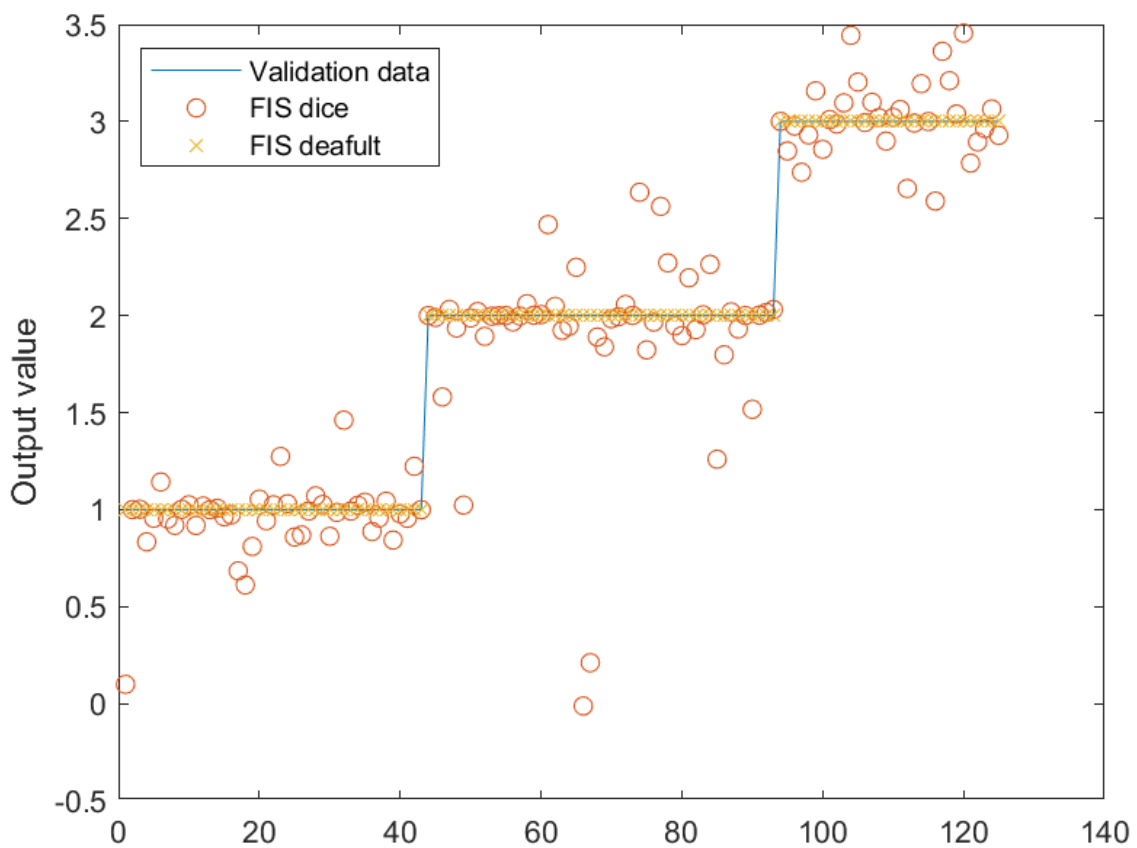
Poziom dopasowania domyślnego FIS dla zestawu testowego: 0,8868

Poziom dopasowania dostrojonego FIS dla zestawu trenującego: 0,9920

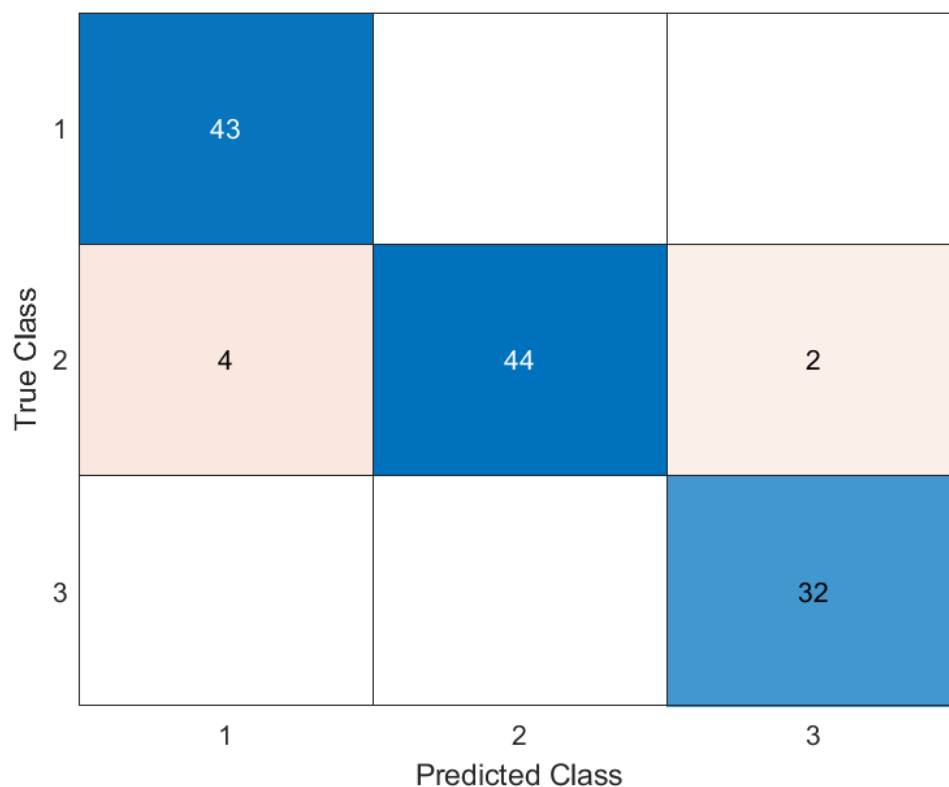
Poziom dopasowania dostrojonego FIS dla zestawu testowego: 0,8113



Rysunek 8 . Wykres najlepszego przystosowania w każdej iteracji.



Rysunek 9. Wykres danych wyjściowych, domyślnego Fuzzy Logic oraz dostrojonego Fuzzy Logic



Rysunek 10. Macierz pomyłek dostrojonego układu FL dla zbioru Wine

Dla zbioru Wine dokładność, precyzja i czułość wynoszą :

$$\text{accuracy} = \frac{43+44+32}{43+4+44+2+32} = 0.952$$

$$\text{precision}(c_1) = \frac{43}{43+4} = 0.915$$

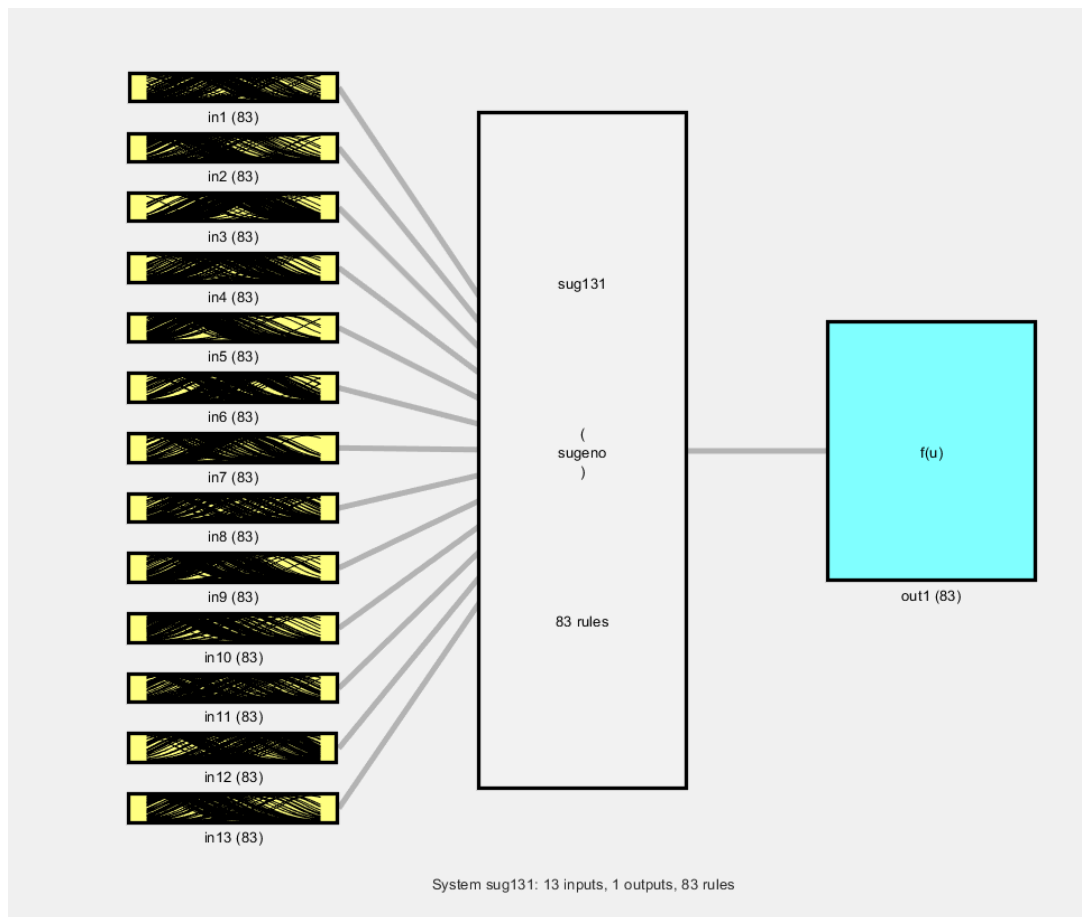
$$\text{precision}(c_2) = \frac{44}{44} = 1$$

$$\text{precision}(c_3) = \frac{32}{32+4} = 0.941$$

$$\text{recall}(c_1) = \frac{43}{43} = 1$$

$$\text{recall}(c_2) = \frac{44}{44+4+2} = 0.880$$

$$\text{recall}(c_3) = \frac{32}{32} = 1$$

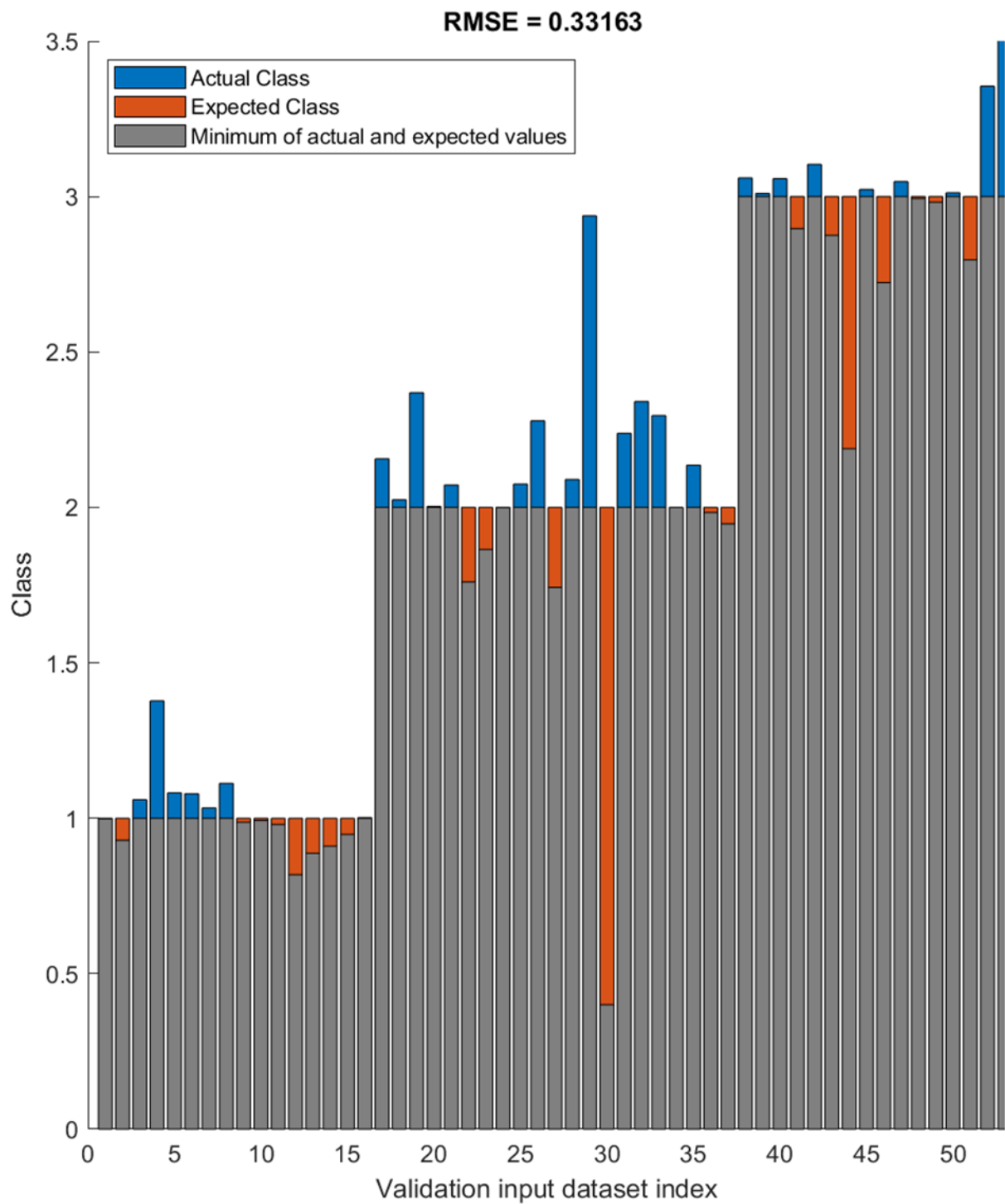


Rysunek 11. Struktura dostrojonego FIS.

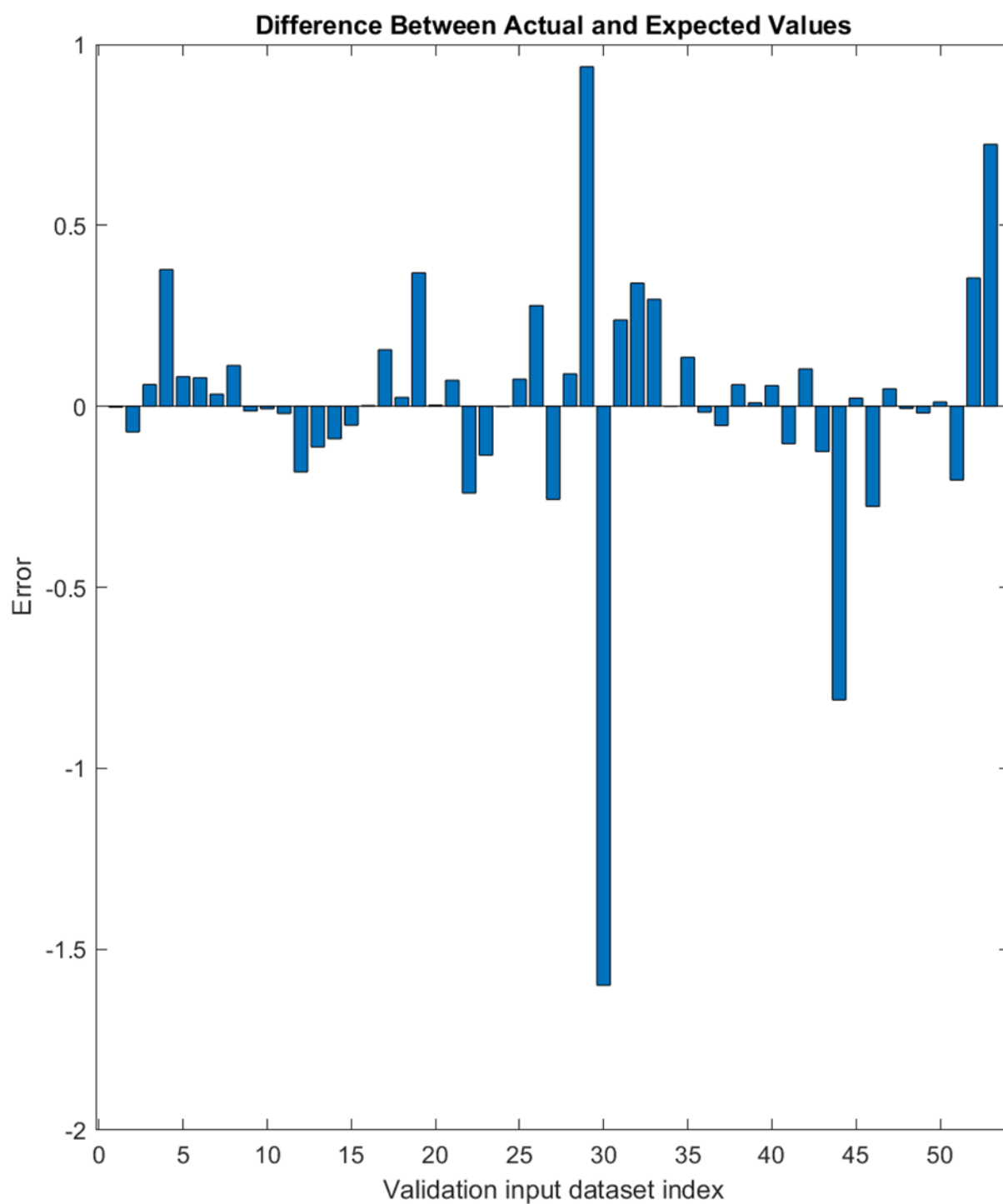
Średniokwadratowy błąd naszego dostrojonego układu Fuzzy Logic dla zestawu WINE wyszedł akceptowalny. Wynik na zbiorze trenującym:

```
>> wykresy
```

```
Training RMSE = 0.327
```



Rysunek 12. Wykres otrzymanego i oczekiwanego wyniku z RMSE na zbiorze testującym



Rysunek 13. Wykres różnicy między otrzymanym a oczekiwanym wynikiem.

4.3 SEEDS

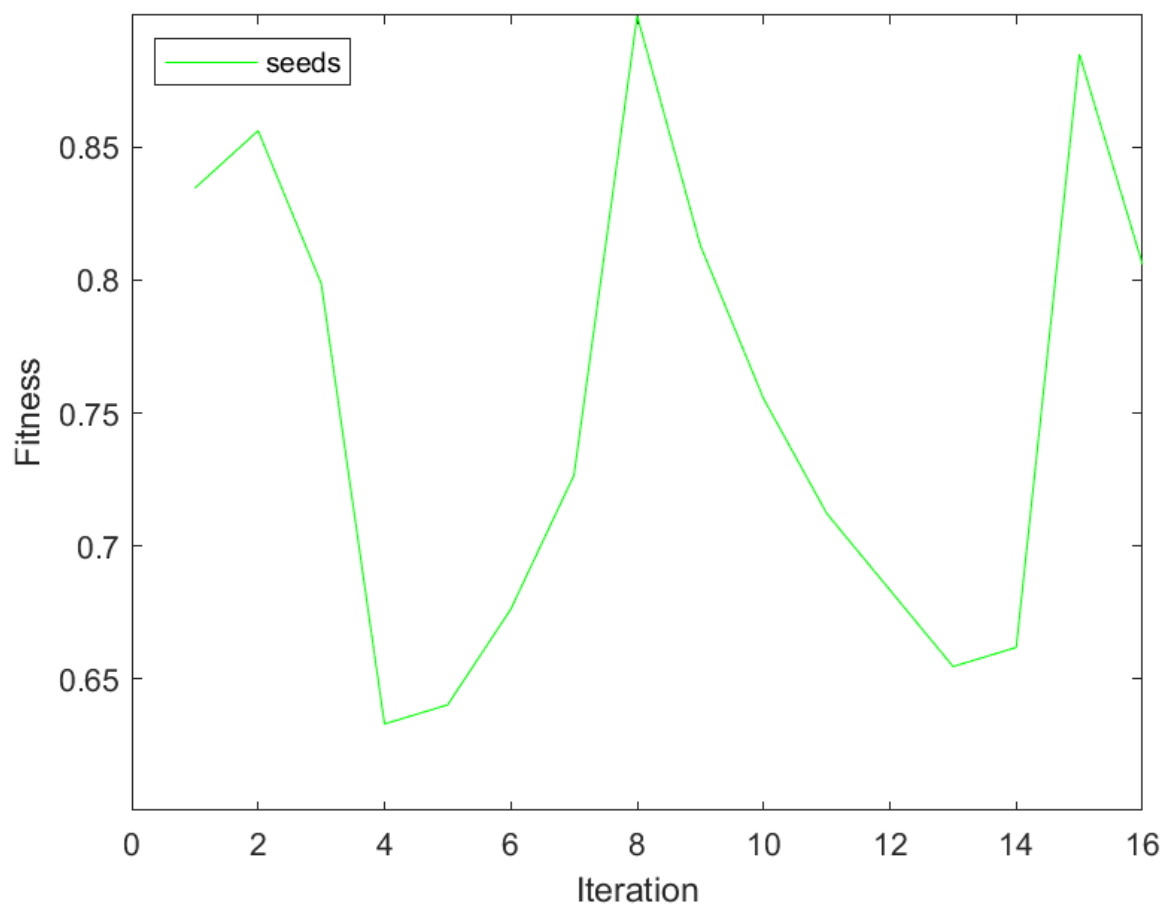
Dla tego zestawu otrzymaliśmy najgorsze wyniki.

Poziom dopasowania domyślnego FIS dla zestawu trenującego: 0,9568

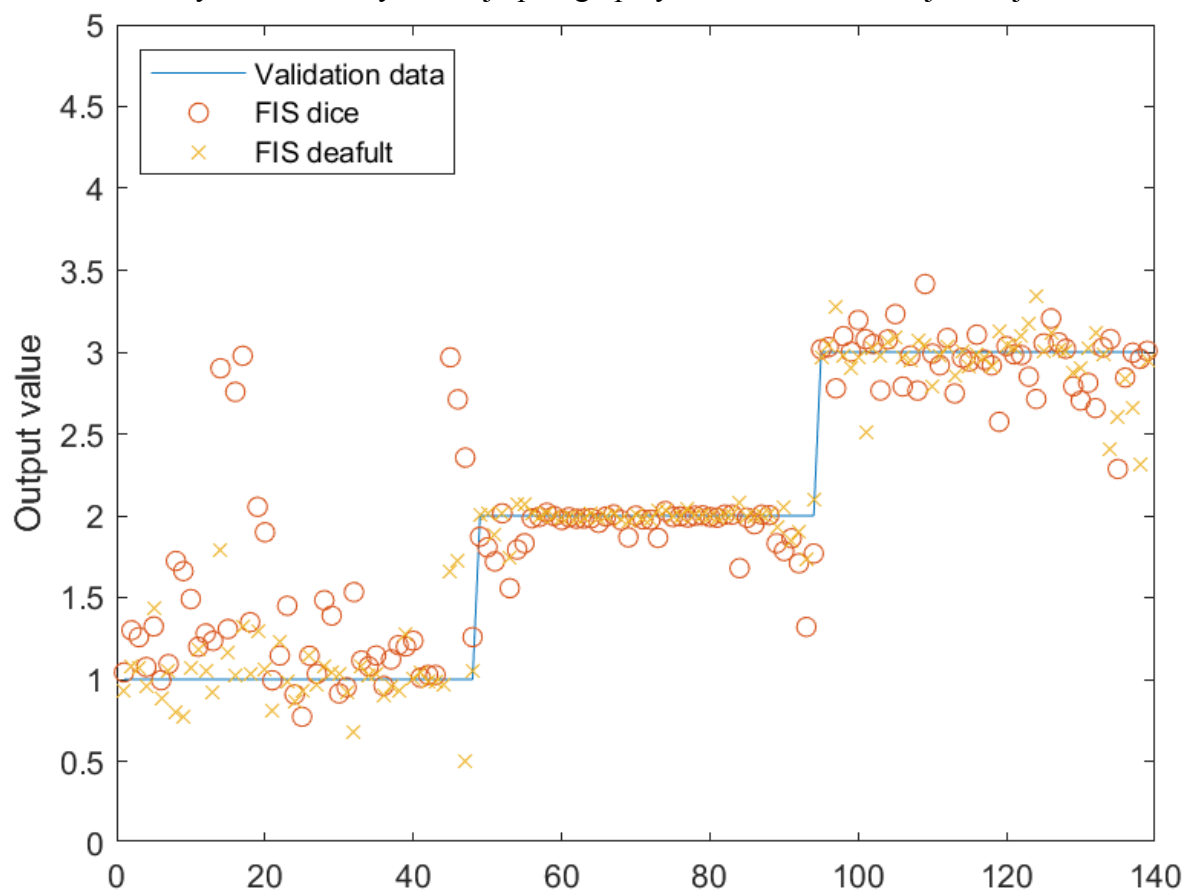
Poziom dopasowania domyślnego FIS dla zestawu testowego: 0,8833

Poziom dopasowania dostrojonego FIS dla zestawu trenującego: 0.8993

Poziom dopasowania dostrojonego FIS dla zestawu testowego: 0,8000



Rysunek 14 . Wykres najlepszego przystosowania w każdej iteracji.



Rysunek 15. Wykres danych wyjściowych, domyślnego Fuzzy Logic oraz dostrojonego Fuzzy Logic

		Predicted Class		
		1	2	3
True Class	1	36	6	6
	2	1	45	
	3		1	44

Rysunek 16. Macierz pomyłek dostrojonego układu FL dla zbioru Seeds

Dla zbioru Seeds dokładność, precyzja i czułość wynoszą :

$$\text{accuracy} = \frac{36+45+44}{36+6+6+1+45+1+44} = 0.899$$

$$\text{precision}(c_1) = \frac{36}{36+1} = 0.973$$

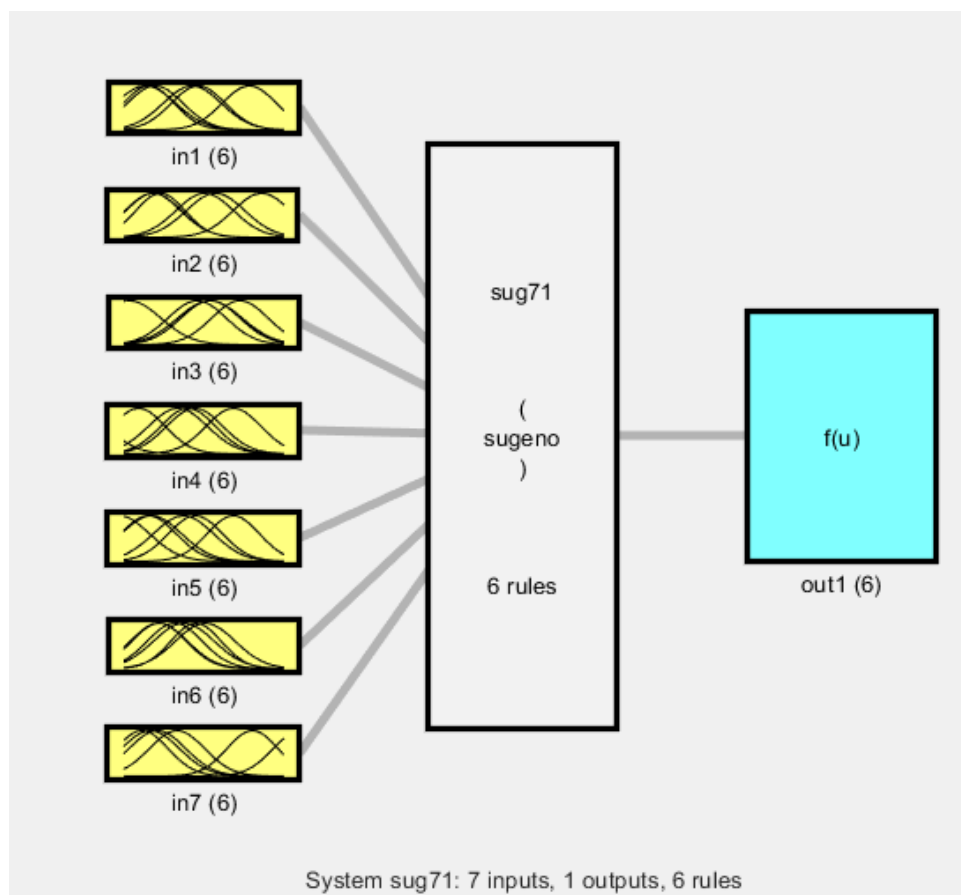
$$\text{precision}(c_2) = \frac{45}{45+6+1} = 0.865$$

$$\text{precision}(c_3) = \frac{44}{44+6} = 0.880$$

$$\text{recall}(c_1) = \frac{36}{36+6+6} = 0.750$$

$$\text{recall}(c_2) = \frac{45}{45+1} = 0.978$$

$$\text{recall}(c_3) = \frac{44}{44+1} = 0.978$$

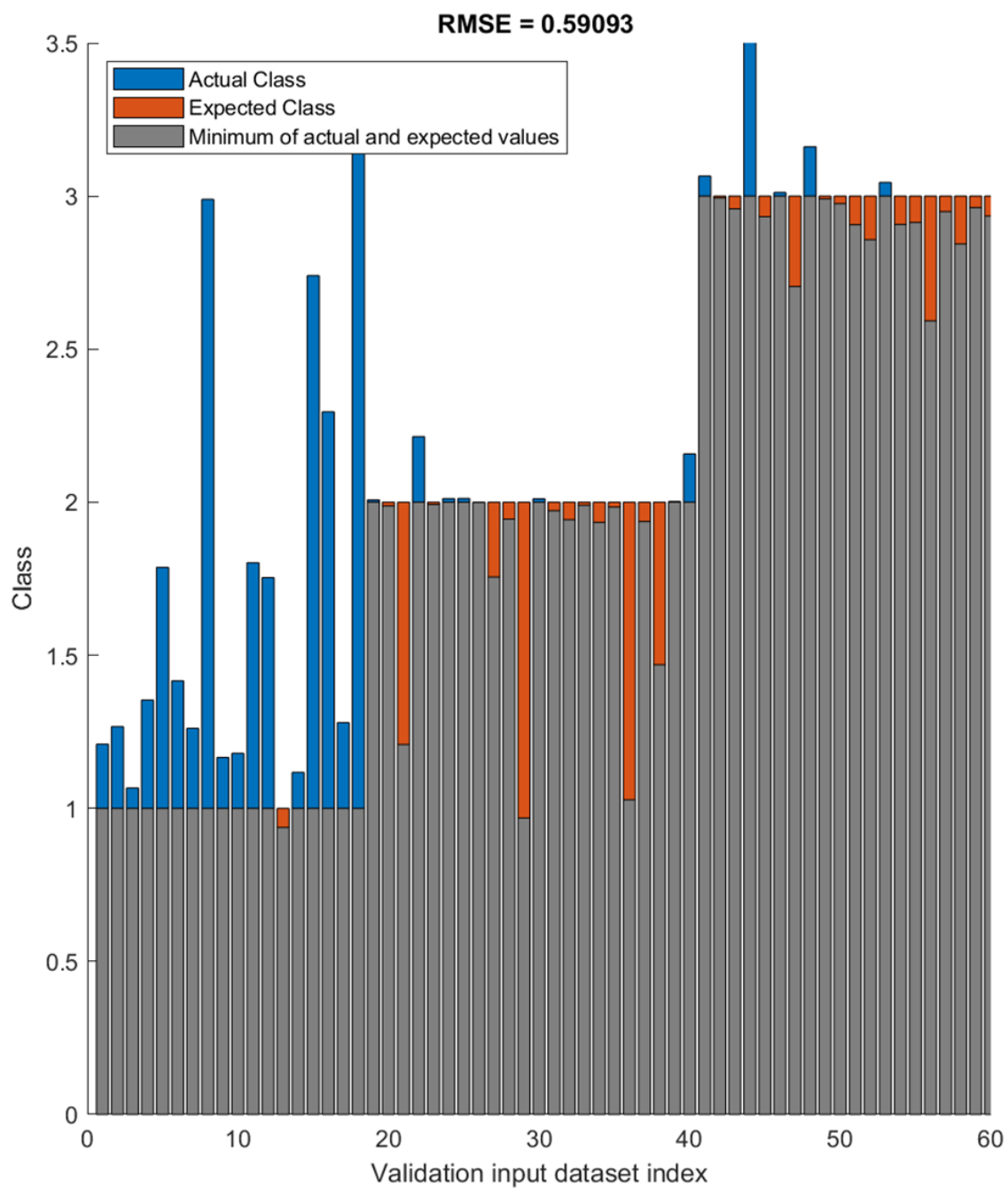


Rysunek 17. Struktura dostrojonego FIS.

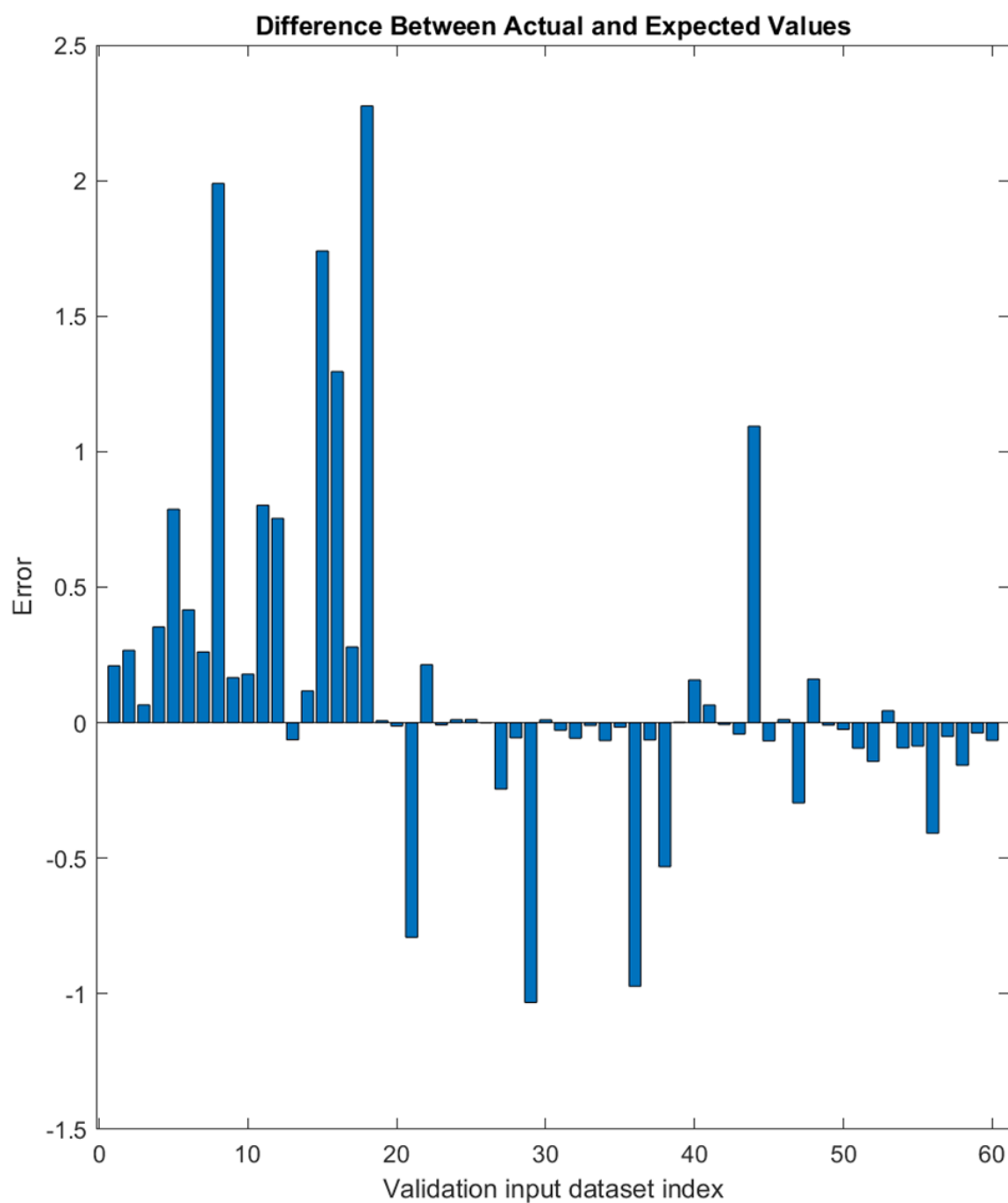
Średniokwadratowy błąd naszego dostrojonego układu Fuzzy Logic dla zestawu SEEDS wyszedł dość duży. Wynik na zbiorze trenującym:

```
>> wykresy
```

```
Training RMSE = 0.535
```



Rysunek 18. Wykres otrzymanego i oczekiwanego wyniku z RMSE na zbiorze testującym



Rysunek 19. Wykres różnicy między otrzymanym a oczekiwanym wynikiem.

Wynik walidacji krzyżowej na zbiorze Iris:

Domyślny FIS: 0,9524

Dostrojony FIS: 0,8952

Wynik walidacji krzyżowej na zbiorze Wine:

Domyślny FIS: 1

Dostrojony FIS: 0,8655

Wynik walidacji krzyżowej na zbiorze Seeds:

Domyślny FIS: 0,6759

Dostrojony FIS: 0,6770

Układ Fuzzy Logic dla każdego zbioru danych znajduje się w repozytorium. Dołączone zostały również tam wygenerowane reguły dla każdego zbioru.

5. Podsumowanie

Znaczny wpływ na otrzymywane wyniki miało odpowiednie dopasowanie parametrów: POPULATION, STALL_MAX oraz M_STALL_MAX oraz liczba wejść zbiorów.

Najlepiej parametry udało dopasować się do zbioru danych Iris. Odpowiednie dostosowanie parametrów dawało zadowalające wyniki już w początkowych iteracjach. Poziom dopasowania domyślnego oraz dostrojonego FIS jest niemal identyczny (a nawet lepszy w przypadku dostrojonego FIS). Wartości otrzymane na podstawie macierzy pomyłek wyszły idealne. Niewielki błąd średniokwadratowy dla tego zestawu również wskazuje na to, że parametry zostały dopasowane odpowiednio i rozwiązanie wygenerowane przy ich pomocy jest zadowalające.

W przypadku zbioru Wine wyniki również były zadowalające, jednak generowały się przez bardzo długi czas (24 godziny), co pozwala nam stwierdzić, że metoda przez nas wykorzystywana nie jest wydajna przy danych zawierających dużą ilość cech (w tym przypadku 13). W przypadku poziomego dopasowania domyślnego oraz dostrojonego FIS widzimy, że FIS domyślny działa lepiej niż dostrojony, jednak różnice są niewielkie. Wartości otrzymane na podstawie macierzy pomyłek wyszły bardzo dobre.

Wartości najgorsze otrzymaliśmy dla zestawu Seeds. Różnice między poziomem dopasowania domyślnego i dostrojonego FIS zwiększają się i możemy zauważyć, że nawet domyślne FIS działa w sposób znacznie gorszy niż dla poprzednich zestawów. Najlepsze przystosowania w kolejnych iteracjach osiągają skrajne wartości, a wartości otrzymane na podstawie macierzy pomyłek uległy pogorszeniu w zestawieniu z poprzednimi zbiorami. Błąd średniokwadratowy również znacząco wzrósł.

W przypadku walidacji krzyżowej CV-5 wyniki zarówno na zbiorze Iris jak i Wine wyszły zadowalające, natomiast w przypadku Seeds były one niskie (jednak w przypadku Iris oraz Wine były one gorsze dla FIS dostrojonego, a dla Seeds - przeciwnie).

Problem z wartościami dla zbioru Seeds może wynikać z nieodpowiedniego dobrania parametrów dla dużej liczby wejść.

Algorytm DGO przynosi względnie akceptowalne wyniki, jednak jego wybór do dostrajania układu Fuzzy Logic nie jest optymalnym podejściem ze względu na czas wykonywania i istnienie skuteczniejszych metod do rozwiązywania tego typu problemów.

6. Bibliografia

Dehghani, M. and Montazeri, Z. and Malik, O.P. and Al-Haddad, K. and Guerrero, J.M. and Dhiman, G., A New Methodology Called Dice Game Optimizer for Capacitor Placement in Distribution Systems (February 19, 2020). Electrical Engineering & Electromechanics, 0(1), 61–64, 2020. doi:10.20998/2074-272x.2020.1.10, Available at SSRN:

<https://ssrn.com/abstract=3761136>

https://pl.wikipedia.org/wiki/Tablica_pomy%C5%82ek?fbclid=IwAR0bXiKjwy1G-YLzDbN6vfZaHn1c_0WVxvIXE-i6f4D8q9HTc2M0_3IoRgU

https://home.agh.edu.pl/~pszwed/wiki/lib/exe/fetch.php?media=med%3Amed-w04.pdf&fbclid=IwAR2jeNI5dfk2tV1kmHXBV769n-Zn4GLJQEW-jQV_vs7OMzYLA2QRu9QbQvQ