

폐지 클러스터링기반 뉴로-퍼지를 이용한 간경화 예측 성능 비교

정은서¹, 노유민², 임도예³, 기희수⁴, 곽근창⁵, 염찬욱⁶

^{1,2,3,4}조선대학교 의생명과학과 학부생

⁵조선대학교 전자공학부 교수

⁶조선대학교 AI융합대학사업단 연구교수

jes020607@chosun.kr, dbalsl025@chosun.kr, doye8208@chosun.kr, ks281537@chosun.kr,

kwak@chosun.ac.kr, walt18@chosun.ac.kr

Performance Comparison of Cirrhosis Prediction Model Using FCM Clustering-based Neuro-Fuzzy

Eun-Seo Jung¹, Yu-Min Roh², Do-Ye Lim³, Hui-Su Ki⁴, Keun-Chang Kawk⁵,
Chan-Uk Yeom⁶

^{1,2,3,4}Dept. of Biomedical Sciences, ChoSun University

⁵Professor, Dept. of Electronic Engineering, ChoSun University

⁶Research Professor, AI Convergence College, ChoSun University

요 약

본 연구는 간경화 예측을 위한 다양한 머신러닝 모델의 성능을 비교하여 최적의 예측 모델을 찾는 것을 목표로 한다. Kaggle의 "Cirrhosis Prediction Dataset"을 활용해 Random Forest 알고리즘으로 중요한 변수 4개(N_Days, Platelets, Prothrombin, Albumin)를 선정하고, grid-ANFIS(grid partitioning-Adaptive Neuro Fuzzy Inference System), subtractive-ANFIS(subtractive-Adaptive Neuro Fuzzy Inference System), FCM(Fuzzy Cluster means) 모델을 평가하였다. 평가 지표로 RMSE를 사용하였으며, grid-ANFIS 모델이 가장 우수한 성능을 보였다. 본 연구는 간경화 예측 및 진단의 정확성을 높이는 데 기여할 수 있을 것으로 기대된다.

1. 서론

간 경화(Cirrhosis)는 간의 만성적 손상으로 인해 간 조직이 섬유화되고, 정상적인 간 기능이 저하되는 질환으로, 전 세계적으로 높은 이환율과 사망률을 보인다. 조기 진단과 예측은 환자의 생존율을 높이고 치료의 성공률을 증대시키는 데 필수적이나, 간 경화의 복잡한 병리학적 특성 때문에 기존 진단 방법들은 예측 정확도에 한계를 보인다[1]. 이러한 이유로 다양한 예측 모델들이 제안되었다.

그러나 환자의 생리적 지표와 관련된 정보들을 활용하는 분석에는 대부분 통계분석(DeLong test, variance analysis 등)이 사용되며 인공지능을 활용하는 경우는 드물다[2]. 인공지능을 이용한 경우 대체로 의료 영상을 기반으로 하는 진단, 예후, 예측에 관련된 프로그램이다[3]. 따라서 의료 영상 기반 예측 모델에 생리적 지표 분석이 추가로 이루어진다면 보다 정확한 진단, 예측에 도움이 될 것이다.

본 연구는 환자의 생리적 지표를 활용한 예측 모델들을 비교하여 간 경화 예측에 가장 적합한 모델

을 식별하는 것을 목표로 한다.

2. 간경화 예측 데이터셋 분석과 변수지정

본 연구에서는 Kaggle에서 제공하는 "Cirrhosis Prediction Dataset"을 활용하였다[4]. 해당 데이터셋은 간 경화 예측을 위한 다양한 환자 기록을 포함하고 있다. 이 데이터셋은 4,145개의 샘플과 13개의 변수로 구성되어 있다. 각 샘플은 환자의 간경화 상태와 관련된 정보를 담고 있다. 데이터의 형식은 CSV이며, 각 변수는 환자의 생리적 지표와 관련된 정보를 포함한다.

위 데이터셋의 변수를 Random Forest 알고리즘을 통해 중요도가 높은 변수 4개를 선정하였다[5]. 분석 결과, N_Days, Platelets, Prothrombin, Albumin가 간경화 예측에 가장 중요한 변수로 나타났다. 'N_Days'는 환자의 생존 기간을, 'Platelets'와 'Prothrombin'은 혈소판과 프로트롬빈 수치를, 'Albumin'은 혈중 알부민 수치를 나타낸다. 이 변수들을 중심으로 분석하여 모델 성능을 높이는 데 기여 하였다.

3. 실험연구 및 결과

Grid-ANFIS는 일정 간격의 삼각형 소속 함수를 생성해 규칙을 정의하고, Subtractive-ANFIS는 subtractive 클러스터링으로 데이터를 클러스터링하여 소속 함수를 할당한다. FCM은 클러스터 중심을 반복 최적화해 복잡한 패턴을 반영하는 퍼지 시스템을 설계한다. 세 함수를 통해 생성된 모델들의 성능을 최소 훈련 RMSE를 기준으로 평가했다. 각 기법은 데이터의 복잡한 패턴을 반영하는 데 차별화된 방식을 사용하며, 최종적으로 가장 적합한 예측 모델을 식별하기 위한 분석을 수행하였다.

	Grid-ANFIS	Subtractive-ANFIS	FCM-ANFIS
최소 training RMSE	0.138552	0.287717	0.346527

표 1 결과 비교

Grid-ANFIS은 가장 낮은 훈련 RMSE(0.138552)와 MSE(0.019197)를 기록하여, 세 모델 중에서 가장 우수한 성능을 보였다. 실제 출력과 예측 출력 간의 차이가 거의 없으며, 예측 성능이 매우 안정적인 모습을 보였다. 그래프에서 실제 출력과 예측 출력이 거의 일치하여, grid-ANFIS이 간경화 예측에 매우 적합하다는 것을 확인할 수 있다.

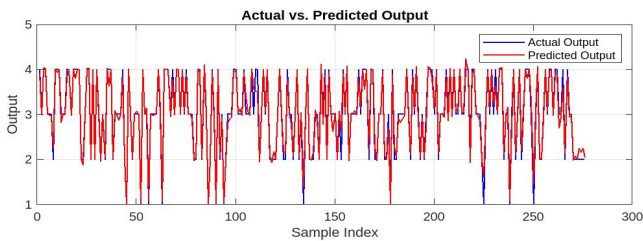


그림 1 Grid-ANFIS 실제 및 예측 출력 비교

Subtractive-ANFIS 모델의 경우, 훈련 RMSE는 0.287717, MSE는 0.082781로 grid-ANFIS에 비해 다소 성능이 저하되었다. 그래프에서 실제 출력과 예측 출력 간의 차이가 약간 더 큰 것을 확인할 수 있다. 이는 grid-ANFIS에 비해 모델의 예측 능력이 떨어짐을 의미한다.

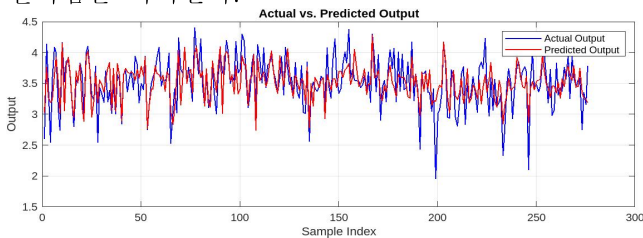


그림 2 Subtractive-ANFIS 실제 및 예측 출력 비교

FCM은 가장 높은 훈련 RMSE(0.346527)와 MSE(0.120081)를 기록하였다. 이 모델의 예측 출력은 실제 출력과의 차이가 가장 크게 나타나, 상대적

으로 예측 정확도가 낮은 모델임을 확인할 수 있다. 그래프에서도 실제 출력과 예측 출력 간의 불일치가 눈에 띄며, grid-ANFIS과 subtractive-ANFIS에 비해 간경화 예측 성능이 떨어짐을 알 수 있다.

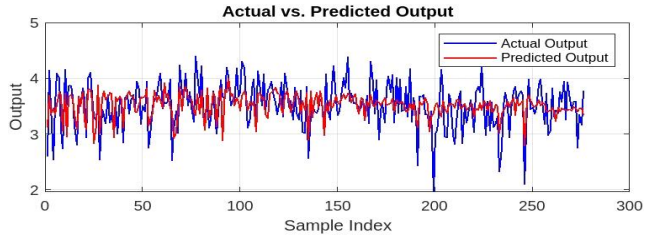


그림 3 FCM 실제 및 예측 출력 비교

연구 결과, 세 가지 FIS(Fuzzy Inference System) 모델 중 Grid-ANFIS가 간경화 예측에서 가장 우수한 성능을 보였다. Subtractive-ANFIS는 다소 성능이 낮았지만 유의미한 결과를 도출했으며, FCM-ANFIS는 가장 낮은 성능을 기록했다.

4. 결론

본 연구는 FCM 기반 퍼지 모델들이 간경화 예측에 효과적임을 확인했으며, 특히 Grid-ANFIS가 최적의 모델로 제안될 수 있음을 시사한다. 향후에는 더 다양한 임상 데이터를 적용하거나 다른 예측 모델과의 결합을 통해 모델을 더욱 고도화함으로써 예측 성능을 높이고, 간경화 예측의 정확성과 신뢰성을 향상시키는 데 중점을 둘 계획이다.

Acknowledgement

이 논문은 2024년도 정부(과학기술정보통신부)의 재원으로 지원을 받아 수행된 AI 헬스케어 융합대학 사업 연구임.

참고문헌

[1] Kim, J., et al. (2020). "Challenges in Liver Cirrhosis Diagnosis and Management: A Review." *Hepatology International*, 14(2), 345-356.
 [2] 김세운, 임형준, 이준영 외 14인, 간경변 환자에서 CTP, MELD, and MELD-Na 점수의 단기 사망률 예측에 대한 비교 분석, 대한소화학회지, 고려대학교 의과대학 내과학교실, Seoul, Korea, 2007, 92-99.
 [3] 노시형, 임동욱, 이충섭 외 2인, 의료영상기반의 간 섬유화 진단을 위한 인공지능 모델 개발, 한국정보처리학회 추계학술대회 자료집, 2022, 462-46.
 [4] Kaggle. (2024). "Cirrhosis Prediction Dataset." Kaggle.
 [5] Breiman, L. (2001). "Random Forests." *Machine Learning*, 45(1), 5-3.