

심근 세포의 전기생리학적 특징을 이용한 인공 신경망 기반 약물의 심장독성 평가

유예담^{1†} · 정다운¹ · Aroli Marcellinus¹ · 임기무^{2*}

¹금오공과대학교 IT융복합공학과, ²금오공과대학교 메디컬 IT융합공학과

An Artificial Neural Network-Based Drug Proarrhythmia Assessment Using Electrophysiological Characteristics of Cardiomyocytes

Yedam Yoo^{1†}, Da Un Jeong¹, Aroli Marcellinus¹ and Ki Moo Lim^{2*}

¹Dept of IT Convergence Engineering, Kumoh National Institute of Technology

²Dept of Medical IT Convergence Engineering, Kumoh National Institute of Technology

(Manuscript received 10 November 2021 ; revised 28 December 2021 ; accepted 29 December 2021)

Abstract: Cardiotoxicity assessment of all drugs has been performed according to the ICH guidelines since 2005. Non-clinical evaluation S7B has focused on the hERG assay, which has a low specificity problem. The comprehensive in vitro proarrhythmia assay (CiPA) project was initiated to correct this problem, which presented a model for classifying the Torsades de pointes (TdP)-induced risk of drugs as biomarkers calculated through an *in silico* ventricular model. In this study, we propose a TdP-induced risk group classifier of artificial neural network (ANN)-based. The model was trained with 12 drugs and tested with 16 drugs. The ANN model was performed according to nine features, seven features, five features as an individual ANN model input, and the model with the highest performance was selected and compared with the classification performance of the qNet input logistic regression model. When the five features model was used, the results were AUC 0.93 in the high-risk group, AUC 0.73 in the intermediate-risk group, and 0.92 in the low-risk group. The model's performance using qNet was lower than the ANN model in the high-risk group by 17.6% and in the low-risk group by 29.5%. This study was able to express performance in the three risk groups, and it is a model that solved the problem of low specificity, which is the problem of hERG assay.

Keywords: Drug safety assessment, Drug cardiovascular toxicity assessment, Artificial neural network, Machine learning, CiPA

1. 서 론

1999년에는 소화 촉진제인 Cisapride가 염전성 심실 빈맥(Torsades de Pontes, TdP)으로 인해 유럽 의약 시장에서

287
퇴출되는 사건이 발생하였다[1]. 이후 이러한 사건들로 인해 모든 약물에 대해 심장부정맥 유도 가능성을 평가하는 가이드라인이 2005년 의약품 규제조화 위원회(The International Council for Harmonisation, ICH)를 통해 제시되었다[2,3]. 의약품 규제조화 위원회는 심실 재분극 지연에 대한 비임상 평가의 S7B와 약물에 의한 심전도 상에서 나타나는 QT간격 연장 및 부작용 유발 작용에 대한 임상평가인 E14에 따라 약물이 심혈관계에 미치는 독성을 평가할 것을 제안하였다. 그 중 비임상 평가 가이드라인 S7B의 내용은 세포의 활동전위 기간 연장에 중요한 영향을 미치는 IKr 채널을 구성하는 유전자 human Ether a-go-go-related gene(hERG)의 이온

*Corresponding Author : Ki Moo Lim
Kumoh National Institute of Technology, Gumi 39177, Republic of Korea

Tel: +82-54-478-7780

E-mail: kmlim@kumoh.ac.kr

†Contributed equally to this work.

본 연구는 금오공과대학교 학술연구비에 의하여 연구된 논문임 (2021).

통로의 막힘 비율을 측정하여 약물의 독성을 평가한다. 이 평가법은 높은 민감도로 고 위험군 약물에 대해 정확한 평가가 이루어지지만 낮은 특이도로 인해 저위험군 약물이 고 위험군으로 평가되기 쉽다[4]. 이것은 약물이 TdP를 유도하지않는 저위험군의 약물들에게 엄격한 규제가 이루어져 약물 개발에 부정적인 영향을 준다[5].

기존의 가이드라인을 개정하는 목표로 2013년 FDA 본사에서 주최한 Think-tank 회의에 13개의 선진 의료기관들이 참석하여 종합 부정맥 시험관내 평가(Comprehensive in vitro Proarrhythmia Assay, CiPA) 프로젝트가 수립되었다[5]. 이 프로젝트는 네 개의 그룹으로 구성되며 그 중 *in silico* working 그룹은 기존의 비임상 가이드라인의 hERG 채널 평가 방식에서 *in silico* 심실모델을 기반으로 다중이온채널을 평가하는 것을 제안한다[6].

이러한 새로운 약물 독성평가 방식이 제안되면서 Dutta 그룹은 심실모델인 O'Hara Rudy 모델의 INa, INaL, ICaL, IKr 채널의 최대전도도를 수정하고 hERG채널의 약물 동적 결합 파라미터가 추가된 모델을 제안하여 *in vitro* 실험의 결과와 *in silico* 모델을 통해 계산된 결과가 유사한 모델을 고안하였으며, *in silico* 모델을 통해 계산되어지는 약물의 TdP 유도 가능성을 나타내는 qNet 지표를 제안하였다[6,7,9]. qNet 지표는 6개 주요한 이온채널(INaL, ICaL, IKr, IKs, IK1, Ito)을 통해 이동하는 이온 전하들의 총합이다.

Li 그룹은 Dutta그룹이 제안한 모델을 통해 계산된 qNet 지표를 odd 논리 회귀 모델의 입력으로 사용하여 약물의 세 가지 위험군(고위험군, 중위험군, 저위험군) 분류를 수행하였다[8]. 세가지 위험군 분류를 위해 고위험군과 중위험군을 구분하는 qNet 임계 값과 중위험군과 저위험군을 구분하는 qNet 임계 값으로 약물의 위험군을 분류하였다[8]. 분류의 정확도는 수신자 조작 특성 곡선(Receiver operating characteristic curve, ROC curve)의 아래면적(Area under the curve; AUC)을 기준으로 고위험군 분류 성능이 0.988, 저위험군 분류 정확도가 0.901이다. hERG 채널의 약물 동적 결합 모델을 추가함으로써 분류의 정확도가 고위험군에서 12.8%, 저위험군에서 4.1% 향상되었지만 수학적 복잡성의 증가라는 단점이 존재한다[8].

Llopis-Lorente 그룹에서는 109개 약물의 TdP 유도군과 TdP 비유도군을 9개의 의사결정나무를 통해 분류하였다[9]. 그들은 *in silico* 심실모델로 Dutta에 의해 수정된 최대 전도도 모델에 약물에 의해 이온채널의 단순 막힘을 구현하여 사용하였다[6]. 분류 모델의 입력으로 사용된 지표는 세 가지로 약물의 농도 특성인 T_x , qNet 지표의 농도의 민감도를 나타내는 T_{qNet} 지표 그리고 APD90과 APD30의 차이인 APD triangulation(APDtri) 지표의 농도에 대한 민감도를 나타내는 T_{triang} 지표이다.

$$T_x = \frac{[D]_{APD \uparrow 10\%}}{EFTPC}$$

$$T_{qNet} = \frac{qNet_{at\ 10 \times EFTPC}}{qNet_{control}}$$

$$T_{triang} = \frac{APDtri_{at\ 10 \times EFTPC}}{APDtri_{control}}$$

$[D]_{APD \uparrow 10\%}$ 는 APD90 (Action potential 90%)이 10% 증가 될 때의 약물 농도이며 EFTPC는 효과적인 유리 치료 혈장농도의 최대값(effective free therapeutic plasma concentration, EFTPC) 농도이다. $qNet_{at\ 10 \times EFTPC}$ 는 EFTPC 농도의 10배일 때 계산된 qNet 값이며 $qNet_{control}$ 은 약물 영향이 없는 상태일때 qNet 값이다. APDtri는 APD90과 APD30의 차이(APD90-APD30)이다. $APDtri_{at\ 10 \times EFTPC}$ 는 EFTPC의 10배농도에서 계산된 APDtri이며 $APDtri_{control}$ 는 약물의 영향이 없는 상태일 때 계산된 APDtri이다. 약물의 위험군의 분류 정확도는 지표에 따라 T_x 를 입력으로 사용한 경우 89.9%, T_{triang} 을 입력으로 사용한 경우 90.8% 그리고 T_{qNet} 은 91.7%을 보였다[9].

Lancaster 그룹은 Ord, tenTusscher, Grandi 심실 모델을 통해 계산된 APD90과 세포내 칼슘 휴지기 농도인 Ca diastolic 지표를 입력으로 하는 주성분 분석(Principal component analysis, PCA)을 통해 약물의 TdP 유도 약물과 TdP 비 유도 약물을 분류하였다. 분류 성능은 수신자 조작 특성의 아래면적(ROC-AUC)이 0.963을 보였으며 비교군으로 제시한 APD90을 입력으로 하는 서포트 벡터 머신(support vector machine, SVM) 모델의 ROC-AUC는 0.854 성능을 보였다[10].

현재까지 수행된 선행연구들의 흐름은 TdP와 상관관계가 매우 높다고 할 수 있는 활동전위 형상 특성 또는 전하특성을 Dutta가 수정한 ORd 심실 모델로부터 도출하고 이것들을 이용하여 논리회귀, 의사결정나무와 같은 선형적 패턴으로 분류하는 고전적인 이진분류법을 사용하였다[11]. 그러나 약물의 독성 영향이 활동전위 특성과 전하특성에서 위험군에 따라 선형적인 분포가 나와야 한다는 보장은 없다. 또한 활동전위 형상과 전하특성 외에도 TdP 유도 약물과 비 유도 약물의 분류에 Ca 농도 형상 또한 높은 상관 관계성이 제시되었다[6,10].

따라서 본 연구에서는 기존 한 개에서 두개의 입력을 사용한 약물의 독성 분류모델의 지표 선정에 대한 불분명함과 분류성능을 개선하기 위해 활동전위 형상 특성, 칼슘농도 형상 특성, 전하특성을 모두 고려한 다중 특성 값들을 입력으로 하는 인공신경망 기반 약물의 위험군을 고위험군, 중위험군, 저위험군으로 분류하는 모델을 제안한다.

II. 연구 방법

본 연구를 위해서 CiPA 프로젝트 그룹에서 업로드한 데이터를 바탕으로 이루어졌으며, 전체 과정은 그림 1과 같다

(github.com/FDA/CiPA). 실험데이터는 미국 Charlese 연구소에서 고 처리량 패치 클램프(High-throughput patch clamp, HTS) 방식으로 측정된 네 가지 종류의 채널(INa, INaL, ICaL, IKr)에 대한 실험 농도(nM)에 따른 채널 막

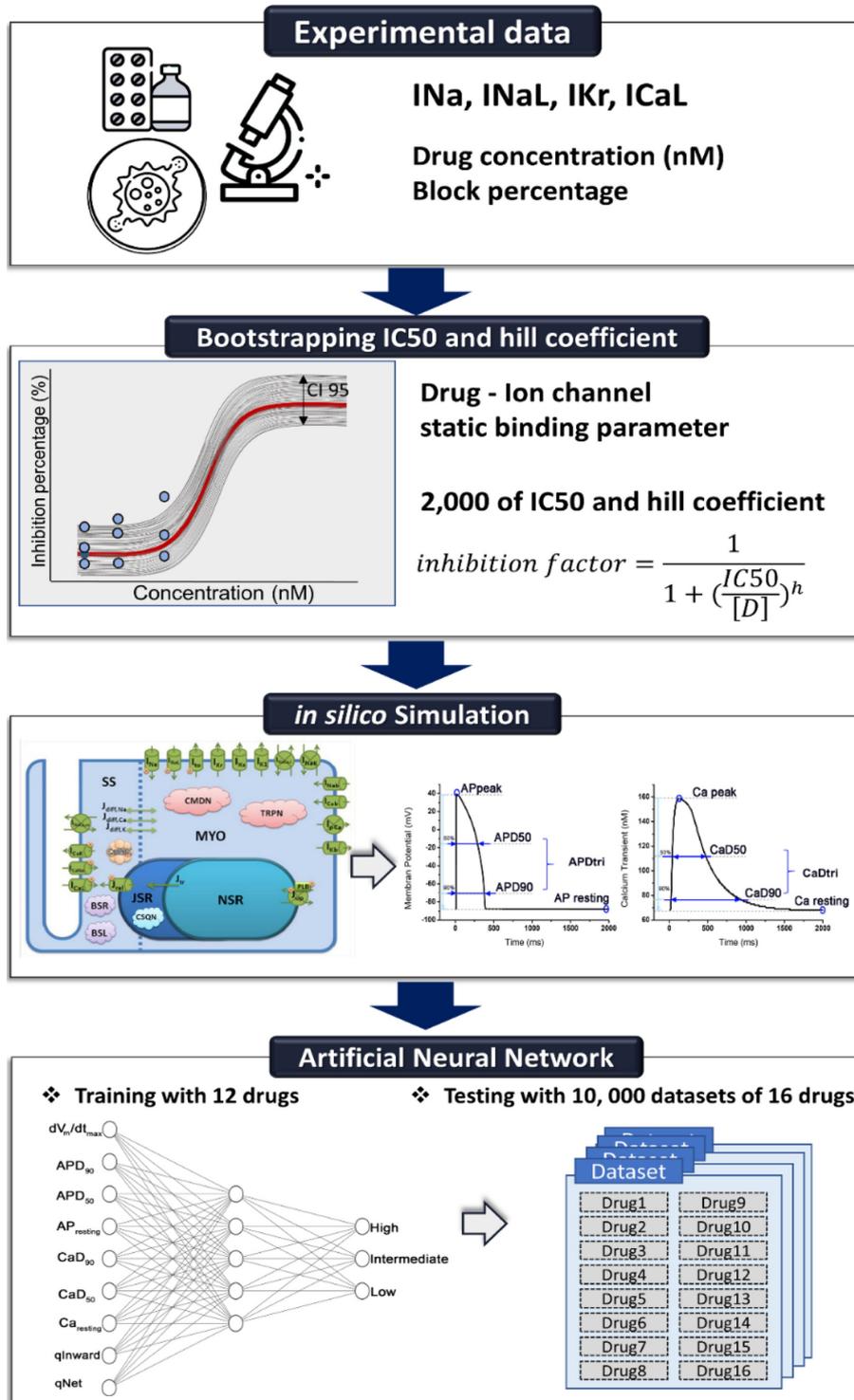


그림 1. 연구진행 과정 모식도
Fig. 1. The schematic diagram of the research process

함비율(block percentage) 정보를 사용하였다[8]. 실험 데이터의 불확실성 정량화를 위해 Chang 그룹이 제안한 Markov-chain Monte Carlo(MCMC) 모델을 통해 부트스트랩 과정을 수행한다[12]. 약물에 따라 네 개 농도 조건에서 시험된 실험데이터를 입력으로 최적화된 Hill 곡선을 최소 제곱법(Least Square Method)을 통해 도출한다. 최적화된 Hill 곡선을 중심으로 신뢰구간 95%이내의 2,000개 Hill 곡선을 유도한다. 2,000개의 Hill 곡선을 통해 반수 저해량(half maximal inhibitory concentration, IC50)와, hill 계수를 추출하여 *in silico* 심실모델의 입력으로 사용한다.

사용한 심근세포 *in silico* 모델은 약물 독성 분류 모델 개발에 사용되어오면서 검증된 Dutta그룹에 의해 수정된 Ord 모델을 사용하였다[6,13]. 이온채널 INa, INaL, INa,k, INaCa, IKr, IKs, IK1, Ito로 이동하는 전류로 인해 발생하는 막전위 차이를 계산하는 모델로서 각 전류식은 Hodgkin-Huxley 가 제안한 식을 기반으로 전개된다(식 2). 각 이온 채널의 약물 영향 반영은 이온 전류식에 이온 전도도 억제 계수 식 (1)를 적용하여 계산되며, 이온 전도도 이온 억제 계수는 부트스트랩을 통해 추출된 2,000개의 IC50와 Hill 계수를 사용하였다. 시뮬레이션 설정으로 약물 농도는 각 약물의 특성 값이라고 할 수 있는 약물 투여 후 최고 혈중 농도(maximum plasma concentration; Cmax)의 1배, 2배, 3배, 4배에 해당하는 농도를 설정하였으며 활동전위 생성을 위한 자극은 2s 간격으로 TdP가 발생하는 조건인 서맥 환경을 구현하였다. 총 1,000번의 자극을 주어 활동전위 시뮬레이션이 정상상태에 도달하도록 설정하였다.

$$\text{inhibition factor(IF)} = \left[\frac{1}{1+(IC50/[D])^h} \right]^{-1} \quad (1)$$

$$I = G(V_m - E) \times IF \quad (2)$$

IC50은 반수 저해량(half maximal inhibitory concentration, IC50) (nM)이며, D는 약물의 농도(nM), h는 Hill 계수이다. I는 각 이온전하에 대한 전류이며, G는 이온채널이 가

지는 최대전도도, V_m은 막전위, E는 평형전위이며 IF는 이온채널 억제계수이다.

우리는 CiPA에서 선택한 약물의 TdP 유도 위험군에 따라 고위험군 8개, 중 위험군 11개, 저위험군 9개인 총 약물 28개로 구성된 약물을 사용하였다[8](표 1).

TdP 예측과 관련 있는 지표 9개를 *in silico* 시뮬레이션을 통해서 계산하며, 이는 활동전위 특성, 칼슘 특성, 이온 전하 특성으로 구성된다. 9개 지표는 약물의 심독성을 나타내는 주요하게 고려되어지는 지표로 선정하였다[6,10]. 활동전위 특성 중 Action potential duration (APD90)은 활동전위 형상에서 최대 진폭으로부터 90% 아래의 탈분극 지점과 재분극 지점 사이의 시간이다. APD50은 활동전위 형상에서 최대 진폭으로부터 50% 아래의 탈분극 지점과 재분극 지점 사이의 시간이다. dV_m/dt_{max}는 활동전위 형상에서 막전위 탈분극 시기의 최대 기울기이며, AP_{resting}은 휴지기 막전위이다. 칼슘 특성으로는 Calcium transient Duration(CaD90)으로 세포내 칼슘의 농도 형상 중 최대 진폭에서 90% 이하 지점 사이의 시간이다. CaD50은 세포내 칼슘의 농도 형상 중 최대 진폭에서 50% 이하 지점 사이의 시간이다. Ca_{resting}, 세포내 칼슘의 휴지기 때의 농도이고, 이온전하 특성의 qNet은 활동전위 기간동안 6개 이온 채널(INaL, ICaL, IKr, IKs, IK1, Ito)을 통해 이동하는 이온전하들의 총량이며, 시간에 따른 전류 변화들의 합으로 그려지는 그래프의 아래면적으로 계산된다[10](식 (3)). qNet은 Li *et al.* 그룹에서 수행한 논리회귀 모델을 이용한 약물의 부정맥 유발 위험도 분류의 입력지표로 사용되었다. qInward는 활동 전위기간 동안 ICaL, INaL 이온채널을 통해 세포내로 향하는 전하들의 약물반응일 때와 정상상태의 비율의 평균이며 식 (4)로 표현된다[12].

$$q_{net} = \int (INaL + ICaL + IKr + IKs + Ik1 + Ito) dt \quad (3)$$

$$q_{inward} = \frac{(ICaL_{drug_AUC}/ICaL_{control_AUC} + INaL_{drug_AUC}/INaL_{control_AUC})}{2} \quad (4)$$

290

표 1. 인공신경망 모델의 훈련과 테스트로 사용한 28개 약물 목록

Table 1. A list of 28 drugs used for training and testing of artificial neural network models

Used / risk level	High	Intermediate	Low
Training	Quinidine	Cisapride	Verapamil
	Sotalol	Terfenadine	Ranolazine
	Dofetilide	Chlorpromazine	Diltiazem
	Bepridil	Ondansetrom	Mexiletine
Testing	Disopyramide	Clarithromycin	Metoprolol
	Ibutilide	Clozapine	Nifedipine
	Vandetanib	Domperidone	Nitrendipine
	Azimilide	Droperidol	Tamoxifen
		Pimozide	Loratadine
		Risperidone	
		Astemizole	

*_drug_AUC*는 약물 특성이 반영된 상태일때 각 이온 채널의 시간에 따른 전류 변화 그래프 아래 면적, *_control_AUC* 은 정상상태일때 각 이온 채널의 시간에 따른 전류 변화 그래프 아래 면적이다.

지표 계산을 위한 하나의 활동전위 선택 기준은 1,000개의 활동전위 중에서 마지막 250개의 활동전위에 대해 탈분극 또는 재분극이 실패한 것을 제외하고, 재분극 시의 dV_m/dt_{repol} 값이 가장 큰 하나의 활동 전위를 통해서 지표들이 계산된다[12]. 약물의 농도별 (C_{max} , $C_{max} \times 2$, $C_{max} \times 3$, $C_{max} \times 4$)로 *in silico* 시뮬레이션을 수행해서 계산된 지표들의 평균값을 인공신경망 모델의 입력으로 할당하였다(그림 1).

제안하고자 하는 인공신경망 모델은 다양한 지표의 조합으로 테스트하였다. 모델은 3가지로 9개의 지표(dV_m/dt_{max} , $AP_{resting}$, APD_{90} , APD_{50} , $Ca_{resting}$, CaD_{90} , CaD_{50} , $qNet$, $qInward$)를 입력으로 하는 9개 노드를 가지는 입력층과 5개의 노드를 가지는 은닉층, 3개 노드를 가지는 출력층으로 구성된 모델(그림 2(a))과 입력 지표 7개(dV_m/dt_{max} , APD_{90} , APD_{50} , CaD_{90} , CaD_{50} , $qNet$, $qInward$)를 입력으로 하는 7개 노드를 가지는 입력층 와 4개노드를 가지는 은닉층과 3개의 노드를 가지는 출력층으로 구성된 모델(그림 2(b)), 그리고 입력지표 5개(dV_m/dt_{max} , APD_{90} , CaD_{90} , $qNet$, $qInward$)의 입력층 5개 노드와 은닉층 4개 노드, 출력층 3개로 노드로 구성된 모델(그림 2(c))의 성능을 비교하였다. 이 3개 모델의 선정은 중위험군 AUC가 0.7이상의 값을 가지는 조합들 중에서 비교하였다. 모델 학습 중에 각 지표 간의 단위 차이로 발생하는 과대적합 혹은 과소적합을 피하고자 입력 데이터들에 대해 *MinMaxScaler*를 사용하였다. 은닉층의 활성화함수는 'ReLU'를 사용하였으며, 출력층의 활성화함수는 'Softmax'를 사용하였다. 출력층 노드에서는 고위험군, 중위험군, 저 위험군으로 다중 분류하였다. 훈련 중 모델의 최적화를 위해 *Leave one out* 교차검증을 하였다. 이는 훈련약물 12개를

순서대로 하나씩 제외하여 제외된 약물 하나를 검증 데이터로 사용하고 11개 약물 데이터를 학습에 사용하는 검증 방식이다. 오차 함수는 'categorical_crossentropy'를 사용하였으며, *optimizer*는 'adam'을 사용하였다. *Batch* 사이즈는 32로 설정하였으며 *epoch*는 100회로 설정하였다.

In silico 시뮬레이션을 통해서 약물별로 농도별로 평균화된 2,000개의 샘플이 획득되고, 훈련 데이터는 12개 약물에 대해서 계산된 24,000개의 샘플과 테스트를 위한 데이터는 16개 약물로부터 획득된 32,000개 데이터가 사용되었다. 모델의 테스트에 사용한 자료 데이터셋은 16개 약물의 샘플이 하나씩 속하며, 각 약물의 2,000개의 샘플 중 무작위로 하나씩 선정하여 만들어진다. 모델의 테스트는 16개 테스트 약물의 자료를 하나씩 포함하는 10,000개의 자료집합을 이용하여 10,000번을 수행한다[8]. 테스트 결과로는 수신자 조작 특성 곡선(receiver operating characteristic; ROC curve)의 면적에 해당하는 AUC(Area under the curve)를 사용하여 모델의 성능을 평가하였다. 결과 비교는 10,000개의 AUC 결과와 특이도(specificity), 민감도(sensitivity)의 도수 분포 중 결과의 신뢰구간 95% 범위와 중앙 값을 대표 값으로 설정하였다(식 (5,6)).

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{5}$$

$$\text{Specificity} = \frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}} \tag{6}$$

예를 들어 *Positive* 값과 *Negative* 클래스가 있다면, *Positive*를 분류 기준으로 *True Positive*는 분류기를 통해 예측된 값이 *Positive* 과 실제 값이 *Positive*로 같은 경우이며, *False Positive*는 분류기를 통해 예측된 값이 *Positive*이지만 실제 값이 *Negative*인 경우이다. *True Negative*는 예측한 값이 *Negative*이면서 실제 값이 *Negative*이며, *False*

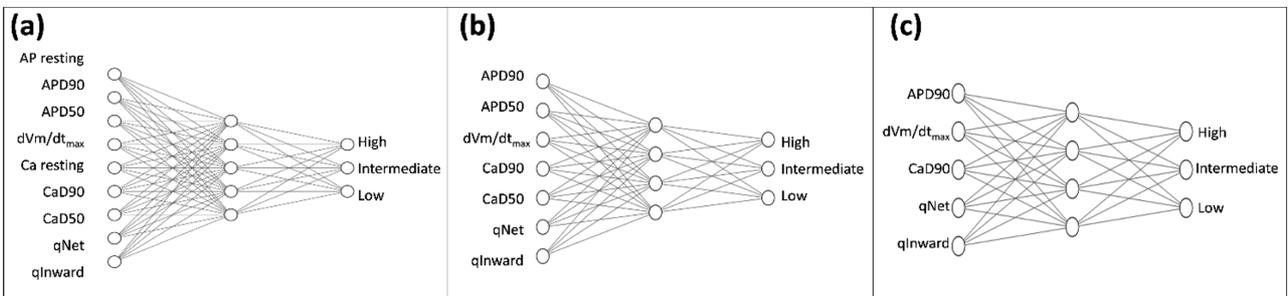


그림 2. 입력지표에 따른 ANN 모델 구조 (a) 9개 입력지표(dV_m/dt_{max} , $AP_{resting}$, APD_{90} , APD_{50} , $Ca_{resting}$, CaD_{90} , CaD_{50} , $qNet$, $qInward$)를 사용한 ANN 모델, (b) 7개 입력지표(dV_m/dt_{max} , APD_{90} , APD_{50} , CaD_{90} , CaD_{50} , $qNet$, $qInward$)를 사용한 ANN 모델, (c) 5개 입력지표(dV_m/dt_{max} , APD_{90} , CaD_{90} , $qNet$, $qInward$)를 사용한 ANN 모델
 Fig. 2. ANN model structure according to input indicators (a) 9 input indicators (dV_m/dt_{max} , $AP_{resting}$, APD_{90} , APD_{50} , $Ca_{resting}$, CaD_{90} , CaD_{50} , $qNet$, $qInward$), and (b) 7 input indicators (dV_m/dt_{max} , APD_{90} , APD_{50} , CaD_{90} , CaD_{50} , $qNet$, $qInward$)

Negative는 분류기로 통해 예측된 값이 Negative이지만 실제 값이 Positive인 값이다. 민감도는 분류기를 통해 예측된 Positive들 중에 분류기가 실제 Positive로 분류한 비율이며, 특이도는 예측된 Negative 중에 실제 Negative의 비율이다.

III. 연구 결과

본 연구에서 입력지표에 따른 약물 TdP 위험군 분류모델의 성능비교를 표 2에 표현하였다. 인공신경망 모델의 성능은 10,000개의 데이터셋을 통해 수행된 테스트 결과의 분포에서 신뢰구간 95% 범위의 중앙 값을 대표 값으로 설정하여 성능을 비교하였다. 테스트 방법은 Li그룹과 동일한 방법을 수행하여 직접적인 비교를 위해 선택되었다. 9개 입력 모델의 성능은 고위험군 AUC는 0.932로 고위험군과 다른 위험군(중위험군과 저위험군)으로 구분하는 성능을 93.2% 나타낸다. 중위험군 AUC는 0.722로 중위험군과 다른 위험군(고위험군과 저위험군)을 구분하는 성능을 72.2% 나타낸다. 그리고 저위험군 AUC는 0.820로 다른 위험군(고위험군과 중위험군)과의 구분하는 성능이 82% 이다. 7개 입력 지표를 입력으로 하는 모델은 고위험군 AUC는 0.931로 고위험군과 다른 위험군(중위험군과 저위험군)으로 구분하는 성능을 93.1% 나타낸다. 중위험군 AUC는 0.759로 중위험군과 다른 위험군(고위험군과 저위험군)을 구분하는 성능을 75.9% 나타낸다. 그리고 저위험군 AUC는 0.880로 다른 위험군(고위험군과 중위험군)과의 구분하는 성능이 88% 이다. 5개 입력지표를 사용한 모델은 고위험군 AUC는 0.932로 고위험군과 다른 위험군(중위험군과 저위험군)으로 구분하는 성능을 93.2% 나타낸다. 중위험군 AUC는 0.722로 중위험군과 다른 위험군(고위험군과 저위험군)을 구분하는 성능을 72.2% 나타낸다. 그리고 저위험군 AUC 0.920로 다른 위험군(고위험군과 중위험군)과의 구분하는 성능이 92% 이다. 고위험군 AUC는 9개 지표를 사용한 모델과 5개 지표를 사용한 모델의 대표 값이 0.932로

같으며, 7개 지표를 사용한 모델과 차이는 0.1%로 더 높다. 중위험군 AUC는 또한 9개 지표 모델과 5개 지표모델의 성능이 0.722로 같지만 7개 지표의 모델과 3.7% 차이로 낮다. 저위험군 AUC는 5개 입력 지표를 사용한 모델이 가장 높으며 9개 지표를 사용한 모델과 10% 차이가 나며, 7개 지표를 사용한 모델과 4% 차이로 높다. 동일한 대표 값을 갖는 9개 지표를 사용한 모델과 5개 지표를 사용한 모델의 고위험군과 중위험군의 95% 신뢰구간을 비교하면 고위험군 AUC의 신뢰구간이 9개 지표를 사용한 모델에서 0.296으로 범위가 5개 지표를 사용한 모델의 신뢰구간의 범위 0.045과 비교할 때 0.251 차이로 5개 지표를 사용한 모델의 신뢰구간이 더 좁다(표 2). 중위험군 AUC는 9개 지표를 사용한 모델의 신뢰구간 95% 범위가 0.204이며 5개 지표를 사용한 모델의 신뢰구간 범위는 0.111로 두 모델의 차이는 0.093으로 5개 모델을 사용한 모델의 신뢰구간이 더 좁다. 민감도와 특이도는 5개 지표를 사용한 모델과 7개 지표를 사용한 모델의 대표 값이 동일하며 9개 지표를 사용한 모델은 다른 두 모델에 비해 특이도는 1.7%, 민감도는 1% 낮은 성능을 보였다(표 2).

입력지표에 따라서 3개의 모델의 AUC 성능을 비교하여 3개 위험군에서 고위험군과 저위험군이 0.9이상의 높은 성능을 가진 5개 입력지표 모델을 대표 모델로 선정하여 Li 그룹에서 수행한 qNet 입력의 논리회귀 모델의 성능과 비교하였다. 두 모델에서 사용한 지표는 채널에 대한 inhibition factor를 사용하여 동일한 *in silico* 모델을 통해 계산되었다. 두 모델의 성능은 민감도는 Li 그룹 모델이 0.59, ANN 모델이 0.84 이며 특이도는 Li 그룹 모델이 0.58이고 ANN 모델이 0.83이다(그림 3). 고위험군 AUC는 Li 그룹 모델이 0.756 저위험군 AUC가 0.625이며 ANN 모델과 비교하였을 때 고위험군 AUC는 5개 입력을 사용한 ANN 모델이 17.6%, 저위험군 AUC가 29.5% 높았다(그림 3).

292

표 2. 약물 독성 분류 ANN 모델의 입력지표에 따른 성능 비교. AUC; Area under the curve of ROC, 9 features (dVm/dtmax, APresting, APD90, APD50, Caresting, CaD90, CaD50, qNet, qInward), 7 features (dVm/dtmax, APD90, APD50, CaD90, CaD50, qNet, qInward), 5 features (dVm/dtmax, APD90, CaD90, qNet, qInward)

Table 2. Comparison of performance according to input features of drug toxicity classification ANN models. AUC; Area under the curve of ROC, 9 features (dVm/dtmax, APresting, APD90, APD50, Caresting, CaD90, CaD50, qNet, qInward), 7 features (dVm/dtmax, APD90, APD50, CaD90, CaD50, qNet, qInward), 5 features (dVm/dtmax, APD90, CaD90, qNet, qInward)

Number of ANN input feature	AUC of High	AUC of Intermediate	AUC of Low	Specificity	Sensitivity
9 features	0.932(0.681-0.977)	0.722(0.629-0.833)	0.820(0.7-0.92)	0.811(0.741-0.872)	0.830(0.748-0.881)
7 features	0.931(0.704-0.977)	0.759(0.685-0.833)	0.880(0.78-0.96)	0.828(0.764-0.865)	0.840(0.768-0.870)
5 features	0.932(0.909-0.954)	0.722(0.667-0.778)	0.920(0.879-1)	0.828(0.795-0.858)	0.840(0.809-0.870)

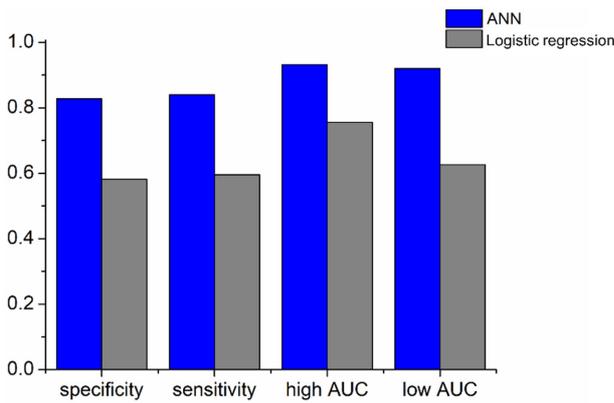


그림 3. Li그룹에서 제시한 qNet을 입력으로 사용한 논리회귀모델의 분류 성능 [8]과 본 연구에서 제시한 5개 입력지표 모델의 분류 성능 비교 AUC; Area under the curve of ROC, ANN; Artificial neural network as input 5 features, Logistic regression; Logistic regression model using qNet as input

Fig. 3. Comparison of classification performance between the logistic regression model using qNet presented in Li group [8] and the model using five input indicators presented in this study. AUC; Area the curve of ROC, ANN; Artificial neural network as input 5 features, Logistic regression; Logistic regression model using qNet as input

IV. 고찰 및 결론

본 연구에서는 수정된 ORd 모델을 통해 계산된 활동전위 형상 특성(APD90, APD50, dV/dtmax, APresting), 칼슘농도 형상 특성(CaD90, CaD50, Caresting), 전하특성(qNet, qInward)을 고려한 다중 특성 값들을 입력으로 하는 인공지능 기반 약물의 TdP 유도 위험성 분류기를 개발하였다. 5개 입력지표(dVm/dtmax, APD90, CaD90, qNet, qInward)를 사용한 모델이 고위험군과 저위험군 분류에서 가장 높은 성능을 보였으며, 7개 입력지표를 사용한 모델이 중위험군 분류에 가장 높은 성능을 보였다.

5개 입력지표의 ANN 모델과 같은 *in silico* 모델과 데이터를 사용한 Li 그룹과 같은 방식인 10,000개의 데이터셋으로 성능 비교를 했을 때 [11], Li 그룹에서 제시한 분류 성능 AUC값 보다 고위험군에서 17.6%, 저위험군에서 29.5% 더 높았으며 특이도와 민감도는 25% 높았다 [8]. 또한 Li 그룹에서 사용한 논리회귀 분류법 대신에 인공지능을 사용함으로써 고위험군, 저위험군 뿐만 아니라 중간위험군 분류도 명시적으로 할 수 있었다.

Li그룹과 Šimundić이 제안한 진단 정확도의 기준으로 성능을 평가할 경우 AUC의 값은 고위험군과 저위험군은 0.9 이상의 'excellent' 정확도 범위이며, 중위험군은 0.7 이상으로 Li그룹에서는 '최소 수용가능 성능', Šimundić의 기준에

서는 '보통' 정확도 범위이다 [14,15]. 결과를 통해 ANN 모델은 기존 ICH의 S7B 가이드라인의 hERG assay 평가의 문제인 낮은 특이성 문제를 해결한 모델이다.

본연구의 한계점으로 독성평가 분류기의 입력 값으로 5가지 파라미터들이 제공되어야 하는 것으로 이것은 선행 연구들에서 필요로 하는 입력 값이 1개인 것에 비하면 매우 많은 편이다. 5개의 신뢰성 있는 파라미터를 획득하기 위해서는 심근세포의 생리/약리 *In silico* 모델의 충분한 신뢰성이 뒷받침되어야 할 것이다. 두번째 한계점으로, 선행연구에서 제시한 qNet, qInward, TqNet 등과 같이 생리/약리적으로 의미 있는 파라미터들의 임계 값을 기준으로 분류한 것에 비해서 본 연구에서 제안한 인공지능망은 분류를 위한 명시적인 임계 값을 제시하지 못한다는 것이다. 이것은 신약개발 연구자가 심장독성평가를 위해 본 인공지능망 분류기를 사용했을 때, 분류된 결과에 대한 타당성을 인과적으로 평가하기가 어렵다는 의미이다. 위 두가지 한계점은 인공지능망 기반의 방법을 사용한다면 필연적으로 만날 수밖에 없다. 그럼 불구하고, 선행연구에 비해 더 높은 독성평가 분류 성능을 가지는 알고리즘을 개발했다는 것은 신약개발 연구분야에서 매우 의미 있는 일이다.

References

- [1] Roden DM, "Cellular basis of drug-induced torsades de pointes", *Br. J. Pharmacol.*, 2008;154:1502-1507, doi:10.1038/bjp.2008.238.
- [2] Shah RR, "Drugs, QTc interval prolongation and final ICH E14 guideline: An important milestone with challenges ahead", *Drug Saf.*, 2005;28(11):1009-1028, doi:10.2165/00002018-200528110-00003.
- [3] Cavero I, Crumb W, "ICH S7B draft guideline on the non-clinical strategy for testing delayed cardiac repolarisation risk of drugs: A critical analysis", *Expert Opin. Drug Saf.*, 2005;4(3):509-530, doi:10.1517/14740338.4.3.509.
- [4] Strauss DG, et al., "Comprehensive In Vitro Proarrhythmia Assay (CiPA) Update from a Cardiac Safety Research Consortium / Health and Environmental Sciences Institute / FDA Meeting", *Ther. Innov. Regul. Sci.*, 2019;53(4):519-525, doi:10.1177/2168479018795117.
- [5] Colatsky T, et al., "The Comprehensive in Vitro Proarrhythmia Assay (CiPA) initiative - Update on progress", *J. Pharmacol. Toxicol. Methods*, 2016;81:15-20, doi:10.1016/j.vascn.2016.06.002.
- [6] Dutta S, et al., "Optimization of an in silico cardiac cell model for proarrhythmia risk assessment", *Front. Physiol.*, 2017;8:1-15, doi:10.3389/fphys.2017.00616.
- [7] O'Hara T, Virág L, Varró A, Rudy Y, "Simulation of the undiseased human cardiac ventricular action potential: Model formulation and experimental validation", *PLoS Comput. Biol.*, 2011;7(5), doi:10.1371/journal.pcbi.1002061.
- [8] Li Z, et al., "Assessment of an In Silico Mechanistic Model for Proarrhythmia Risk Prediction Under the CiPA Initiative", *Clin. Pharmacol. Ther.*, 2019;105(2):466-475, doi:

- 10.1002/cpt.1184.
- [9] Llopis-Lorente J, Gomis-Tena J, Cano J, Romero L, Saiz J, Trenor B, “InSilico Classifiers for the Assessment of Drug Proarrhythmicity”, *J. Chem. Inf. Model.*, 2020;60(10):5172-5187, doi:10.1021/acs.jcim.0c00201.
- [10] Lancaster MC, Sobie EA, “Improved Prediction of Drug-Induced Torsades de Pointes Through Simulations of Dynamics and Machine Learning Algorithms”, *Clin. Pharmacol. Ther.*, 2016; 100(4):371-379, doi:10.1002/cpt.367.
- [11] Ismail WN, Hassan MM, Alsalamah HA, Fortino G, “CNN-based health model for regular health factors analysis in internet-of-medical things environment”, *IEEE Access*, 2020;8:52541-52549, doi:10.1109/ACCESS.2020.2980938.
- [12] Chang KC, *et al.*, “Uncertainty quantification reveals the importance of data variability and experimental design considerations for in silico proarrhythmia risk assessment”, *Front. Physiol.*, 2017;8:1-17, doi:10.3389/fphys.2017.00917.
- [13] Hwang M, Lim C-H, Leem CH, Shim EB, “In silico models for evaluating proarrhythmic risk of drugs”, *APL Bioeng.*, 2020;4(2):021502, doi:10.1063/1.5132618.
- [14] Simundic AM, “Measures of Diagnostic Accuracy: Basic Definitions”, *Ejifcc*, 2009;19(4);203-211.
- [15] Li Z, *et al.*, “General Principles for the Validation of Proarrhythmia Risk Prediction Models: An Extension of the CiPA In Silico Strategy”, *Clin. Pharmacol. Ther.*, 2020;107(1):102-111, doi:10.1002/cpt.1647.