

## 원자력발전소 유량제어계통에서 머신러닝 기반 신호예측에 관한 연구

고진욱\*,\*\* · 김홍집\*\*†

### A Study on Machine Learning Based Signal Prediction in Flow Control Systems of Nuclear Power Plants

Jin Wook Ko\*,\*\*, Hong Jip Kim\*\*†

Key Words : Valve Flow Coefficient(밸브유량계수), Random Forest(랜덤포레스트), Multi-Layer Perceptron(다층퍼셉트론),  
Data Base(데이터베이스), Machine Learning(기계학습)

#### ABSTRACT

This study proposes a machine learning-based approach for predicting lost signals specifically, flow rate( $Q$ ) and pressure( $P$ ) in flow control system by utilizing the inherent characteristics of control valves, particularly the flow coefficient( $C_v$ ). Experimental data were obtained by measuring system flow rate( $Q$ ) and pressure( $P$ ) corresponding to various valve opening rates( $LVDT$ ). Using the acquired dataset, correlations among variables were analyzed, and signal recovery models developed using Random Forest( $RF$ ) and Multi-Layer Perceptron( $MLP$ ) algorithms. The model evaluation results showed high prediction performance, with the coefficient of determination( $R^2$ ) reaching up to 0.96. The proposed method demonstrates potential for enhancing the operational safety and reliability of flow control systems.

#### 1. 서론

유량제어계통에서 효율적인 운영을 위해서는 각 계통의 정확한 데이터(압력, 유량 등)가 필수적이며, 데이터 상실은 스마트 플랜트 운영에 큰 위험요소로 작용할 수 있다. 특히 계통의 운전 데이터에 의존적으로 작동되는 제어밸브는 유량 및 압력 등을 조절하는 유량제어계통에서의 핵심요소로서, 그 고유특성을 활용하면 상실된 신호를 복원할 수 있다. 특히 원자력발전소에서 사용되는 공압구동제어밸브는 안전등급기기로 안정적인 작동이 보장되어야 하며, 제어신호 상실 시 후쿠시마 사고 등과 같은 엄청난 재해의 원인이 되기 때문에 즉각적인 조치를 통해 사고를 방지하는 역할을 한다.

기존 연구에서 제어밸브 유량계수( $C_v$ ) 및 압력손실계수( $K$ )를 전산유체역학(CFD)을 통해 도출하는 방식이 주로 사용되었으나, 실시간 적용이 어려운 한계가 있다<sup>(1)</sup>.

특히 원자력발전소에서 사용되는 계통의 시뮬레이션을 위해서 전산유체역학을 기본으로 실험데이터와 해석 데이터를 병행해서 사용하는 방식이 주로 사용되어 왔다<sup>(2)</sup>. 또한, 1차원적인 프로그램으로는 AMESIM, SimulationX, FlowMaster 등이 있으나, 시스템의 1차원적인 시뮬레이션을 통하여 예측하는 방법론을 적용함에 따라 그 범위가 매우 한정적이며, 특히 밸브 작동에 대한 복잡한 메커니즘의 구현이 어려워 예측한 시뮬레이션 결과와 실험을 통해 얻은 결과와 차이를 가진다.

김재형 등은 제어밸브의 일반적인 타입인 글로브 밸브의 유량과 압력 사이의 관계를 규정하는 유량계수( $C_v$ )를 전산유체역학을 활용한 해석적인 방법으로 접근하여 분석하였다<sup>(3)</sup>. 하지만 국부적인 시뮬레이션 결과를 기반으로 하기 때문에 유량계수와 유량, 압력 사이의 관계에 대해서 한계가 존재한다. 또한 밸브의 유체제어능력을 향상시키기 위한 계통의 유동장 설계와 관련된 연구는 주로 전산유체역학(CFD)을 이용

\* 국립한밭대학교 소프트웨어중심대학사업단(National Program of Excellence in Software, Hanbat National University)

\*\* 충남대학교 기계공학부(School of Mechanical Engineering, Chungnam National University)

† 교신저자, E-mail : khongjip@cnu.ac.kr

한 실험계획법(ROD)을 통해 이루어지고 있다<sup>(4)(5)</sup>.

이러한 단점들을 극복하고자 본 연구에서는 ASME Class 300 lb, 2 inch 글로브타입 공압구동제어밸브를 대상으로 유량제어계통에서의 밸브개폐율(LVDT), 밸브 전·후단 압력( $P_1$ ,  $P_2$ )와 밸브후단유량( $Q$ )에 대하여 실험을 통해 데이터를 취득하고, 취득된 데이터를 통해서 밸브유량계수( $C_v$ ), 압력손실계수( $K$ )를 계산, 각 데이터 간의 상관관계를 탐색하였다. 각 데이터간의 상관관계를 고려하여 랜덤포레스트(RF) 및 다층퍼셉트론(MLP, Multi-Layer Perceptron) 모델을 선택하였다. 랜덤포레스트(RF)는 비선형 관계를 효과적으로 처리하고, 이상치에 대한 강건성 및 변수 중요도 분석이 용이하다는 장점을 가진 앙상블 모델로, 계측 신호의 노이즈와 일부 누락이 발생하는 실제 산업 환경에 적합하며, 다층퍼셉트론(MLP)은 복잡한 비선형 함수 근사에 뛰어난 성능을 보이는 심층 신경망 구조로, 변수 간 상관관계가 고차원적일 경우 정밀한 예측성능을 기대할 수 있다. 따라서, 본 연구에서는 정확도와 일반화 성능을 동시에 고려하고, 실험측정값(압력, 유량)의 물리적 특성에 따라 모델을 분리 적용하는 모델을 제안하고자 한다.

## 2. 데이터베이스 구축

### 2.1 데이터 취득 실험

학습을 위한 데이터셋 구축을 위한 실험장비는 Fig. 1과 같이 제어를 위한 공압구동밸브와 계통운전정보 취득을 위한 부대설비로 구분된다. 공압구동밸브는 공기의 압력에 의한 힘을 이용하여 스프링 복원력을 활용하여 운전된다. 본 실험에 사용된 공압밸브는 ASME Class 300 lb, 2 inch 글로브 타입으로 공급압력 상실 시 스프링 복원력에 의해 개방되는 형태인 Fail-to-Open 형태의 안전기능을 수행하는 모델을 적용하였다. 데이터 취득 시 사용되는 부대설비는 유체를 저장하는 고압저장압력용기( $V$ ), 유체가 이송되는 배관 및 Table 1과 같은 운전정보를 취득할 수 있는 Data Acquisition System(DAQ)을 포함한다.

실험은 총 10개의 운전 시나리오를 기반으로 수행되었으며, 밸브의 개폐율(LVDT)은 10~90%까지 불규칙하게 제어 하면서 데이터를 1초 간격으로 연속적으로 수집하였다. 그 결과 총 277,720개의 데이터를 확보하였으며, 주요 측정 변수인 밸브 개폐율(LVDT)은 10~90% 사이, 밸브 전후단 압력( $P_1$ ,  $P_2$ )은 약 0.1~4.0 bar 사이, 밸브후단유량( $Q$ )은 약 1~115 LPM 사이의 값을 측정하였다. 이러한 측정 범위는 실제 유량제어조건을 충실히 반영하였으며, 모델 학습의 효율성을 고려하여 정규화(Normalization)하였다.

Table 1 Data Description

Symbol	Description	Specification (S/N)
$V_1$	Pressure Vessel (3 m <sup>3</sup> , 10 bar)	-
$P_1$	Valve Inlet Pressure, bar	SENSYS (040412D051)
$P_2$	Valve Outlet Pressure, bar	SENSYS (040412D051)
$Q$	Valve Outlet Flowrate, LPM	Dongyang (DY-LWGY-80C)
LVDT	Valve Opening Rates, %	KOMEIN (017082200005)
DAQ	Data Acquisition System, 32 ch.	YOKOGAWA (S5N808730)



Fig. 1 Data acquisition test system

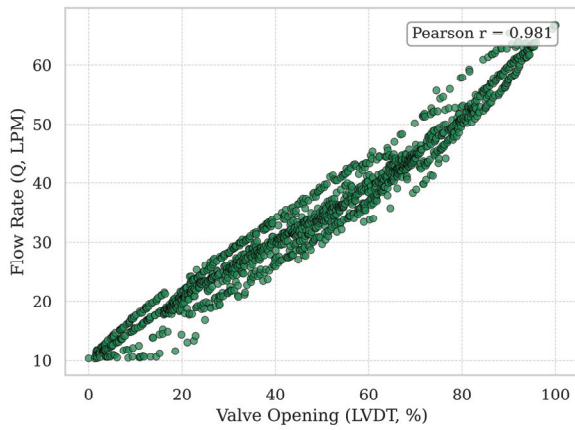
### 2.2 데이터 전처리

본 연구에서는 1초 간격으로 수집된 데이터를 그대로 사용하지 않고, 신호상실 시 공정변수를 통해 제어하고자 하는 변수는 밸브개폐율(LVDT)임에 따라 본 연구에서는 1초 간격으로 수집된 데이터를 그대로 사용하지 않고 그 변화량을 0.1 mm 간격으로 정리하여 각 케이스별 수집된 데이터의 평균을 활용하여 정리하였다. 또한 밸브고유계수를 평가할 때, 식 (1)과 같은 밸브 전·후단 압력차를 사용하였다.

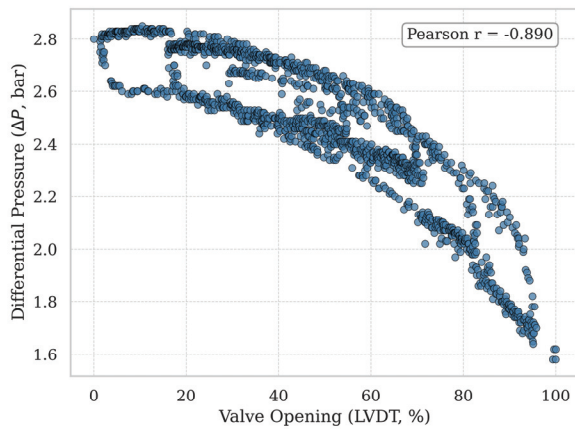
$$\Delta P = P_1 - P_2 \tag{1}$$

밸브개폐율(LVDT)과 차압( $\Delta P$ ), 유량( $Q$ )와의 상관관계 분석을 위해, 식 (2)로 정의되는 피어슨 상관계수(Pearson Correlation Coefficient)를 분석하였다. Mukaka에 의하면 기계학습 및 데이터마이닝 분야에서 피어슨 상관계수( $r$ )를 해석할 때 일반적으로  $|r| < 0.3$ 은 약한 상관,  $0.3 \leq |r| < 0.7$ 은 중간 정도의 상관,  $|r| \geq 0.7$ 은 강한상관으로 분류된다고 제시하였다<sup>(6)</sup>.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \tag{2}$$



(a)



(b)

Fig. 2 Results of Pearson-based correlation analysis

여기서  $x_i$ 는 독립변수의  $i$ 번째 관측값,  $y_i$ 는 종속변수의  $i$ 번째 관측값,  $\bar{x}$ ,  $\bar{y}$ 는 각 값의 평균값,  $n$ 은 총 관측값의 수를 의미한다.

Fig. 2에서 보는 것과 같이 밸브개폐율(LVDT)과 차압( $\Delta P$ )에 대한 관계는  $r = -0.89$ 로 강한 음의 상관관계, 밸브개폐율(LVDT)와 유량( $Q$ )에 대한 관계는  $r = 0.981$ 로 강한 양의 상관관계를 보이고 있음을 확인할 수 있다. 하지만 센서 오차, 신호누락, 노이즈 등의 문제를 포함하고 있어 이를 바로 학습모델에 적용할 경우 학습 성능에 악영향을 미칠 수 있다. 따라서, 모델의 성능향상 및 일반화 능력을 확보하고, 왜곡된 통계값을 보정 및 해석의 정확성 향상을 위해 이상치를 제거할 필요가 있다.

### 2.3 IQR 기반 이상치 탐색

센서오차, 신호누락, 노이즈 등의 이상치를 찾아내는 통계적 방법으로 IQR(Interquartile Range, 사분위수 범위)을 사용하여 이상치를 제거하였다. IQR은 데이터의 중앙 집중

경향과 분산을 고려하여 이상치를 탐지하는 비모수적 방법으로 특히, 정규분포를 가정하지 않아도 되며, 데이터가 편중되어 있는 경우, 강건성(Robustness)이 뛰어난 장점을 가지고 있어, 극단값에 민감한 평균이 표준편차 기반 방법보다, 이상치 탐지에서 더 안정적인 결과를 제공할 수 있다<sup>(7)</sup>.

IQR은 제1사분위수( $Q_1$ )와 제3사분위수( $Q_3$ )의 차이로 정의된다. 여기서 제1사분위수( $Q_1$ )은 전체 데이터 중 하위 25%에 해당하는 값이고,  $Q_3$ 는 전체 데이터 중 상위 75%에 해당하는 값이다. 이때 IQR은 중앙 50% 데이터 범위를 나타내며, 데이터의 분산 정도를 파악할 수 있다. 이상치 판단 기준은 식 (3), (4)과 같으며, 경계값 배수로는 1.5를 사용한다. 1.5는 약한 이상치(mild outlier)를 식별하기 위한 통계적 분포에서 나온 경험적 기준으로 정규분포 또는 유사한 분포를 가정할 때, 데이터의 약 99.3%가 IQR의 1.5배 내에 포함된다는 성질에서 유도되었다<sup>(8)</sup>.

$$Lower\ Bound = Q_1 - 1.5 \times IQR \quad (3)$$

$$Upper\ Bound = Q_3 + 1.5 \times IQR \quad (4)$$

Fig. 3과 같이 이상치 제거 후 피어슨 상관계수가 소폭 감소한 것을 확인할 수 있는데, 이는 일부 극단값이 변수 간의 선형 관계를 인위적으로 강화했을 가능성을 시사할 수 있다. 이는 이상치 제거를 통해 보다 정제된 데이터로 신뢰할 수 있는 상관관계를 분석할 수 있었다고 해석될 수 있다.

이외에도 Isolation Forest, Local Outlier Factor 등 머신러닝 기반 방법론이 존재하지만, 이상치에 대한 직관적인 설명이 어려운 단점을 가지고 있기 때문에 일반적으로 통계학적인 방법론을 적용한 후 필요에 의해 머신러닝 기반 방법론이 적용된다.

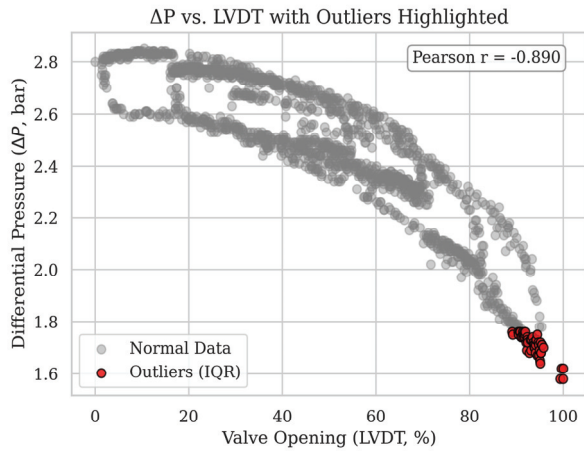
## 2.4 밸브고유계수

### 2.4.1 밸브유량계수( $C_v$ )

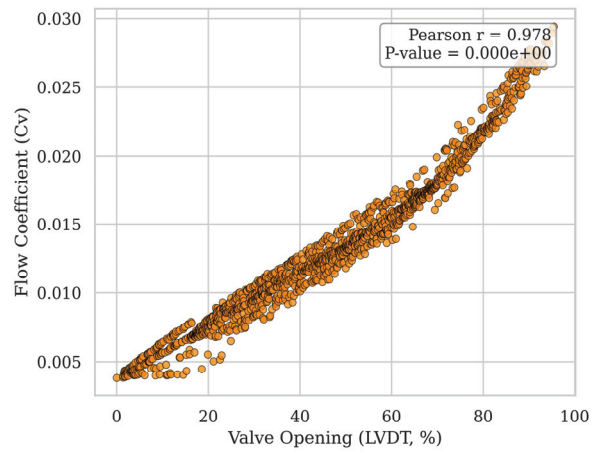
밸브유량계수( $C_v$ )는 밸브를 통과하는 유체의 유량과 압력 강하 간의 관계를 정량화한 지표로, 밸브의 유동특성을 대표하는 고유계수이다. 일반적으로 밸브유량계수( $C_v$ )는 밸브개도율(LVDT)에 따라 변화하며, 특정 구조를 갖는 밸브에 대해서는 실험 데이터를 통해 계산된다. 본 연구에서 밸브유량계수( $C_v$ )는 예측된 압력 및 유량의 예측신호의 정확도를 평가하기 위해 활용된다. 밸브유량계수( $C_v$ )는 다음 식 (5)과 같이 계산된다.

$$C_v = 1.17 Q \sqrt{\frac{S}{\Delta P}} \quad (5)$$

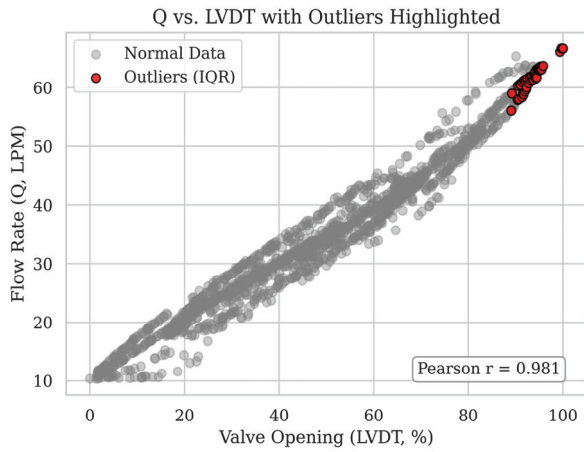
여기서  $Q$ 는 유량( $m^3/h$ ),  $S$ 는 유체의 비중,  $\Delta P$ 는 밸브의 전·후단 차압( $Pa$ )을 나타내며, 상기 변수들은 단위가 SI 단



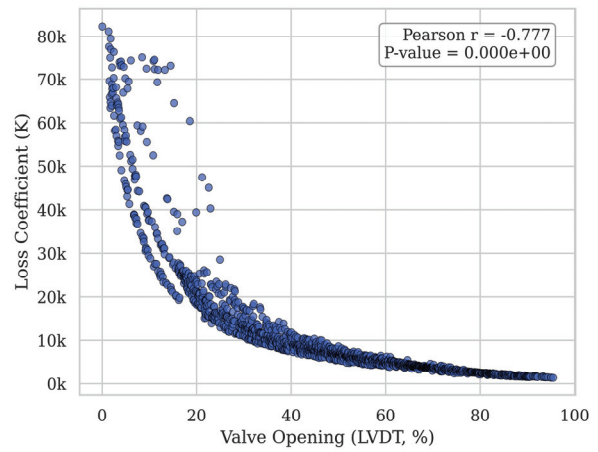
(a)



(a)



(b)



(b)

Fig. 3 Data distribution after outlier removal using IQR

Fig. 4 Valve characteristic coefficient curve

위임에 따라 보정상수 1.17을 곱하여 계산하였다. Fig. 4 (a)와 같이 계산결과 피어슨 상관계수( $r$ )은 0.978로 매우 높은 상관관계를 보였으며,  $p$ -value는 0으로 상관관계가 통계적으로 매우 유의미한 것을 확인할 수 있다.

#### 2.4.2 압력손실계수( $K$ )

압력손실계수( $K$ )는 유체가 밸브를 통과할 때 발생하는 압력 손실의 정도를 나타내는 무차원 계수로 유체 흐름의 에너지 손실을 정량화하여 밸브의 저항특성을 평가하는 주로 사용된다. 본 연구에서는 압력에 대한 예측 정확도를 평가하기 위해 사용되며, 다음 식 (6)과 같이 계산된다.

$$K = \frac{\Delta P}{\frac{1}{2}\rho U^2} \quad (6)$$

여기서  $\rho$ 는 유체의 밀도( $kg/m^3$ ),  $U$ 는 유속( $m/s$ )으로 정의된다. Fig. 4와 같이 계산결과 피어슨 상관계수( $r$ )은

0.978로 매우 높은 상관관계를 보였으며,  $p$ -value는 0을 확인할 수 있다. 여기서 사용된  $p$ -value는 관측된 통계량이 귀무가설 하에서 발생확률을 의미한다. 일반적으로 유의확률이라고도 불리며, 0.05보다 작으면 관측된 결과가 우연히 발생할 가능성이 낮다고 판단하여, 귀무가설을 기각하고 대립가설을 채택하게 된다. 즉 상관관계가 우연히 발생했을 가능성이 사실상 없으며, 두 변수 간의 관계가 통계적으로 매우 유의미함을 뒷받침할 수 있는 의미로 해석된다.

### 3. 예측모델 개발

#### 3.1 랜덤포레스트(Random Forest)

랜덤포레스트는 다수의 결정트리(Decision Tree)로 부터 예측된 값의 평균 또는 기중평균을 출력하는 앙상블 기법들 중 하나이다<sup>(9)</sup>.

실험을 통해 취득한 데이터는 IQR을 통해 이상치를 제거

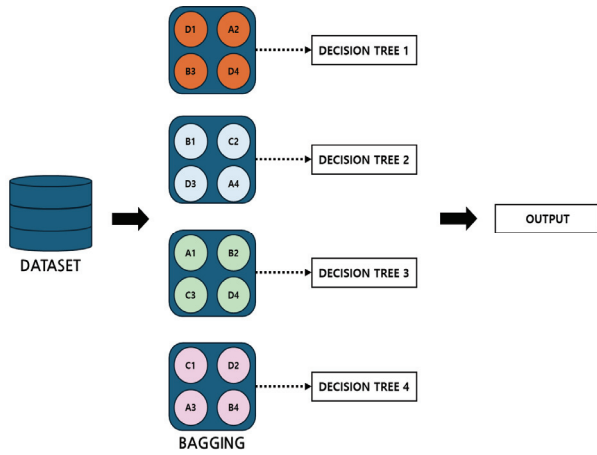


Fig. 5 Bagging process in random forest

했지만, 여전히 일부 이상치가 존재할 가능성과 단일트리의 경우 학습 데이터에 대한 과적합 가능성을 배제할 수 없어 일부 이상치의 영향을 줄이는 특징이 있는 랜덤포레스트 기법을 적용하였다. 랜덤포레스트는 훈련 과정에서 구성된 다수의 결정트리로부터 분류 또는 평균 예측치를 출력하는 배깅(bagging)을 사용하는 앙상블 모형으로 편의(Bias)를 유지하면서 분산을 낮추는 결과를 가져온다<sup>(10)</sup>.

모델의 평가는 평균제곱오차(MSE, Mean Squared Error)와 결정계수( $R^2$ , Coefficient of Determination)의 계산결과를 통해 수행하였다. MSE는 예측값과 실제값 사이의 평균적인 오차를 제공하여 나타낸 값으로 다음 식 (7)과 같이 계산된다.

$$MSE = \frac{1}{n} \sum (y_{true} - y_{pred})^2 \quad (7)$$

여기서  $y_{true}$ 는 실제값(ground truth),  $y_{pred}$ 는 모델의 예측값(predicted value),  $n$ 은 샘플 개수를 의미한다. MSE는 값이 작을수록 모델의 성능이 좋다고 판단할 수 있으나, 제곱을 했기 때문에 이상치에 민감하며, 단위가 원래 데이터의 제곱단위임에 따라 해석이 어려울 수 있는 단점이 존재하기 때문에 일반적으로 결정계수( $R^2$ )를 통해 해석된다.

$R^2$ 는 다음 식 (8)과 같이 계산되며, 모델이 실제 데이터를 얼마나 잘 설명하는 지를 보여주며, 1에 가까울수록 모델이 실제 데이터를 잘 설명한다고 판단한다.

$$R^2 = 1 - \frac{\sum (y_{true} - y_{pred})^2}{\sum (y_{true} - \bar{y})^2} \quad (8)$$

여기서  $y_{true}$ 는 실제값(ground truth),  $y_{pred}$ 는 모델의 예측값(predicted value),  $\bar{y}$ 는 실제 데이터의 평균을 의미한다.

랜덤포레스트 방법을 이용한 예측에 대한 모델 평가결과는 Table 2와 같다. 결정계수( $R^2$ ) 기준 압력손실계수(LVDT)

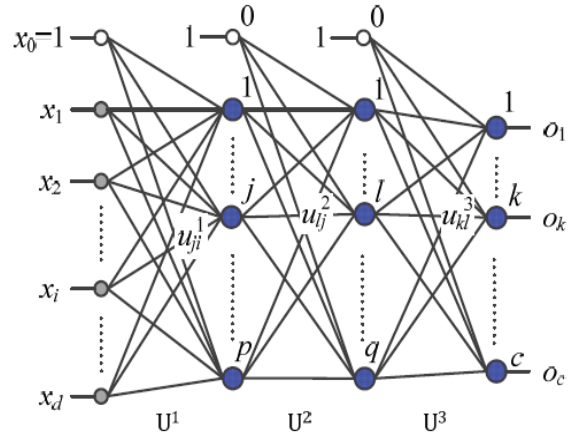


Fig. 6 Structure of 3-layer perceptron

Table 2 Model evaluation for random forest

Input \ Output	LVDT		TRIANGLE		Q	
	MSE	$R^2$	MSE	$R^2$	MSE	$R^2$
LVDT	-	-	7.814	0.843	2.608	0.947
TRIANGLEP	0.005	0.842	-	-	0.012	0.645
Q	7.467	0.952	41.968	0.731	-	-

를 독립변수로 유량(Q)을 예측한 결과가 최대 0.9522로 높은 예측력을 보여주고 있으며, 유량(Q)을 독립변수로 차압(TRIANGLEP)을 예측한 결과가 0.6456으로 가장 낮은 것을 확인할 수 있다. 전반적으로 차압(TRIANGLEP)을 독립변수로 사용한 경우 결정계수( $R^2$ )는 다른 결과에 비해 낮은 예측력을 보여주고 있어, 이에 대한 추가적인 예측 방법론의 적용이 필요하다고 판단하였으며, 이를 위해 다층퍼셉트론(MLP, Multi-Layer Perceptron) 신경망 모델을 통해 예측하였다.

### 3.2 다층퍼셉트론(Multi-Layer Perceptron)

다층퍼셉트론은 여러 개의 퍼셉트론을 결합한 다층구조를 이용하여 선형분리가 불가능한 상황을 해결한다. 다층퍼셉트론의 동작은 다음 식 (9), (10)이 보여주는 바와 같이 벡터  $x$ 를 벡터  $o$ 로 매핑하는 함수로 볼 수 있다. 다음 Fig. 6은 MLP의 매핑과정을 설명한다. 다층퍼셉트론은 여러 퍼셉트론을 결합한 구조인 만큼, 연산도 퍼셉트론의 연산이 연달아 일어나는 방식이다. 현재의 문제는 차압(TRIANGLEP)에 대한 예측정확도를 높이는 것으로 비선형성이 강하기 때문에 2층보다는 3층 퍼셉트론이 더 효과적이다. 하지만 4층 이상의 모델은 과적합 위험이 발생할 가능성이 있다<sup>(11)</sup>.

$$2\text{-Layer Perceptron} : o = f(x) = f_2(f_1(x)) \quad (9)$$

Table 3 Progress model optimization for diff. pressure(*TRIANGLEP*) prediction based on input features(*LVDT*, *Q*)

Input Feature	Random Forest		MLP		MLP + RepeatKFold		MLP + RepeatKFold + Hyperparameter Tuning	
	<i>MSE</i>	<i>R</i> <sup>2</sup>	<i>MSE</i>	<i>R</i> <sup>2</sup>	<i>MSE</i>	<i>R</i> <sup>2</sup>	<i>MSE</i> ± <i>CI</i>	<i>R</i> <sup>2</sup> ± <i>CI</i>
<i>LVDT</i>	7.8143	0.8434	0.0070	0.8749	0.0068	0.8745	0.0063±0.0011	0.8833±0.0120
<i>Q</i>	41.9686	0.7312	0.0137	0.7556	0.0133	0.7544	0.0124±0.0017	0.7708±0.0158

$$3\text{-Layer Perceptron} : o = f(x) = f_3(f_2(f_1(x))) \quad (10)$$

분석 대상 데이터의 스케일링은 입력 데이터의 스케일 차이가 해결, 역전파(Backpropagation) 과정에서 가중치 업데이트 시 입력값이 너무 크거나 작으면 기울기가 급격히 감소하는 경사소실(Vanishing Gradient)의 문제해결을 위해 Min-Max Scaling을 통해 정규화하여 해결한다. 다음 식 (11)은 Min-Max Scaling 과정을 보여준다.

$$X_{std} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (11)$$

여기서,  $X_{std}$ 는 정규화된 데이터,  $X$ 는 원본데이터,  $X_{min}$ ,  $X_{max}$ 는 해당특성(feature)의 최소값과 최대값을 의미한다.

예측결과 다층퍼셉트론 모델이 랜덤포레스트 모델보다 높은 예측력을 가지는 것을 보여준다. 전반적으로 *MSE*는 랜덤포레스트가 더 적은 것을 볼 수 있지만, *R*<sup>2</sup>의 경우 다층퍼셉트론 모델이 더 높은 결과를 가짐에 따라 더 좋은 일반화 성능을 보일 가능성이 크다는 것을 확인할 수 있다. 또한 *MSE*가 더 커진다는 것은 차압(*TRIANGLEP*)과 다른 변수들 간의 비선형적인 관계를 가질 가능성을 예상할 수 있다.

### 3.3 교차검증 기반 모델 일반화 및 최적화

상기 결과에서 *R*<sup>2</sup>와 *MSE*의 차이를 종합적으로 평가해보면 다층퍼셉트론이 더 복잡한 관계를 잘 학습하고 예측하고 있지만, 일부 노이즈까지 학습했을 가능성을 가지고 있고, *MSE*가 랜덤포레스트 모델 대비 다층퍼셉트론 모델에서 크기 때문에 훈련 데이터에 대한 과적합 가능성을 배제할 수 없다.

이러한 과적합을 고려하기 위해서는 K-Fold 교차검증 방법을 이용한다<sup>(12)</sup>. K-Fold 교차검증은 데이터를 K개의 폴드로 나누고, 한 폴드를 테스트 데이터로 사용하고 나머지를 학습데이터로 사용하는 방법이다. 특정 데이터 분할에서 나온 성능 결과에 의존할 가능성을 보완하기 위해 5개의 폴드로 나누어 20%의 검증 데이터와 80%의 훈련데이터를 총 10번 반복하여 예측하였다. 예측결과 *MSE*는 랜덤포레스트 모델을 활용한 결과에 근사하게 접근하였고, *R*<sup>2</sup> 또한 개선된 것을 확인할 수 있다.

K-Fold 교차검증을 통한 결과가 만족스러운 결과임에도 모델의 성능을 최적화하기 위해 하이퍼파라미터 튜닝을 적용

하였다. 다층퍼셉트론의 하이퍼파라미터는 모델의 성능, 학습속도, 일반화 능력에 중요한 영향을 미치기 때문에 적절한 하이퍼파라미터를 찾지 않으면 과소적합, 과대적합 등의 문제가 발생할 수 있다. 따라서 상기 결과에 대한 성능을 높이기 위해 GridSearchCV를 적용하여 하이퍼파라미터를 튜닝하였다.

본 학습에 적용된 grid search는 3개의 은닉층과 2개의 learning rate를 설정하여 각 조합에 대해 개별 모델을 학습시키고 평가하여 가장 성능이 좋은 하이퍼파라미터를 선택하도록 하였다. 이후 최적의 파라미터를 찾은 후 K-Fold 교차검증을 통해 예측하였고, 최종적으로 결과는 Table 5와 같다. 예측결과 최종적으로 *MSE*는 감소하였고, 결정계수(*R*<sup>2</sup>)는 증가하는 결과를 확인할 수 있으며, 이는 예측정확도가 높아졌다고 판단할 수 있다.

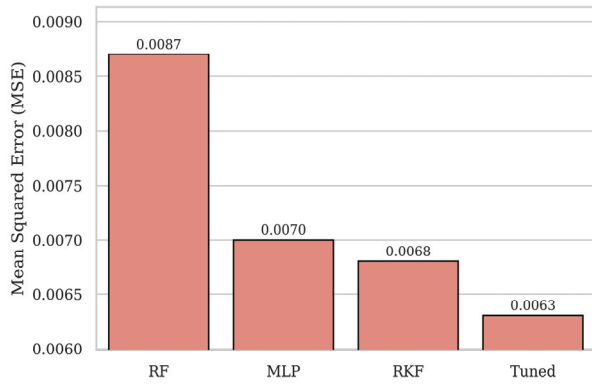
### 3.4 밸브고유계수를 활용한 모델 검증

본 연구에서는 머신러닝 모델을 통해 예측된 밸브개폐율(*LVDT*), 차압(*TRIANGLEP*), 유량(*Q*)를 이용하여 밸브 유량계수( $C_v$ ), 압력손실계수( $K$ )를 계산하고, 이를 원래 실험 데이터를 활용하여 계산된 밸브고유계수( $C_v$ ,  $K$ )와 비교하여 모델의 성능을 평가하였다.

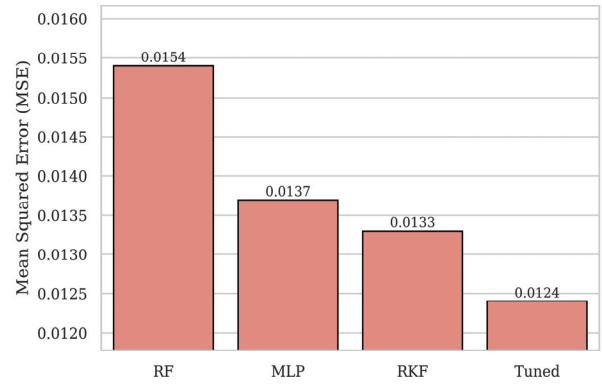
밸브고유계수는 밸브의 형상적 특성을 반영하는 계수로서, 동일한 밸브에서 동일한 조건(유체의 물성, 온도, 압력, 밸브 개폐율 등)이 유지될 경우 본질적으로 변하지 않는 값으로 간주된다. 또한 밸브유량계수( $C_v$ )는 밸브 개폐율에 따라 변하지만 동일한 밸브에서는 같은 조건에서 일정한 값으로 유지된다<sup>(13,14)</sup>. 밸브압력손실계수( $K$ )는 밸브를 통과하는 유체의 에너지 손실을 나타내는 계수로, 유량과 압력차가 변하더라도, 특정 밸브의 형상과 유체 흐름 조건이 일정하면 밸브압력손실계수( $K$ )는 일정하게 유지됨이 전산유체역학(CFD) 및 실험 결과를 통해 입증되어 있다<sup>(15-17)</sup>.

기준값이 되는 밸브고유계수( $C_v$ ,  $K$ )는 최초 데이터셋을 활용하여 계산하였고, 평가를 위한 밸브고유계수( $C_v$ ,  $K$ )의 계산에서는 랜덤포레스트를 통해 예측된 유량( $Q_{predicted}$ )과 다층퍼셉트론을 통해 예측된 압력( $TRIANGLEP_{predicted}$ )을 사용하여 계산하였다.

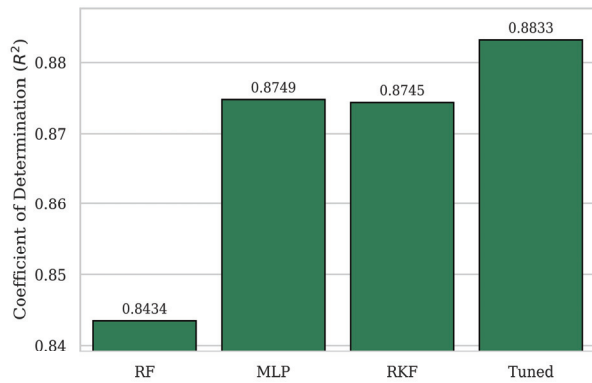
Table 4은 밸브고유계수를 통해 예측된 결과를 보여주고 있다. 밸브유량계수( $C_v$ )의 경우 *MSE*가 0.00161로 예측값과 실제값 간의 평균제곱오차가 매우 낮은 것을 확인할 수 있으



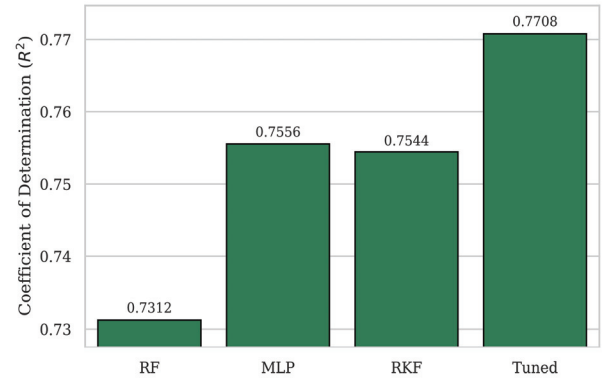
(a)



(a)



(b)



(b)

Fig. 7 Effect of Optimization on predictive performance ( $LVDT$ )

Fig. 8 Effect of Optimization on predictive performance ( $Q$ )

며,  $R^2$ 은 0.9691로 모델이 실제값을 96.91% 설명할 수 있을 정도로 높은 성능을 보여주고 있다.

압력손실계수( $K$ )의 경우  $MSE$ 가 0.00389로 밸브유량계수( $C_v$ )보다는 상대적으로 높은 오차라고 판단할 수 있으나,  $R^2$ 은 모델이 실제값을 88.46% 설명이 가능한 수준으로 전반적으로 높은 수준이라고 판단할 수 있다.

Fig. 7, 8은 예측단계별 성능변화를 보여주고 있다. 여기서  $RF$ 는 랜덤포레스트,  $MLP$ 는 다층퍼셉트론을 통한 최초 예측,  $RKF$ 는 K-Fold 교차검증, 마지막으로  $Tuned$ 는 하이퍼파라미터 튜닝 후 결과를 의미한다. 전반적으로 예측성능은 좋아지는 경향을 확인할 수 있다. 하지만  $MLP$ 에서  $RKF$ 단계로 넘어가면서 결정계수( $R^2$ )값이 소폭 감소한 것을 확인할 수 있는데, 이는 반복 교차검증 과정에서 모델의 일반화 성능이 보다 엄격하게 평가되었기 때문에 발생하는 것으로 예상할 수 있다. 이는 기존 모델이 일부 데이터에 과적합되었을 가능성을 시사하며, 이후 하이퍼파라미터 튜닝을 통해 성능이 다시 향상되는 경향을 확인하였다.

Table 4 Comparison result of  $C_v$  vs.  $K$

Value	$MSE$	$R^2$
$C_{v,actual}$ vs. $C_{v,predicted}$	0.00161	0.9691
$K_{actual}$ vs. $K_{predicted}$	0.00389	0.8846

#### 4. 결 론

본 연구에서는 제어밸브의 유량계수( $C_v$ )와 압력손실계수( $K$ )를 활용하여 계통신호( $LVDT$ ,  $TRIANGLEP$ ,  $Q$ )를 예측하는 머신러닝 기반 접근법을 제안하였으며, 이를 통해 신호 상실 시 계통 신호 복원이 가능함을 실험적으로 검증하였다. 연구의 주요 결과를 요약하면 다음과 같다.

- 1) 기존 연구에서는 제어밸브의 유량계수 및 압력손실계수를 전산유체역학 또는 실험적 방식으로 도출하는 접근이 주를 이루었으나, 본 연구에서는 머신러닝 기반 모델을 활용하여 계통신호 복원을 가증하게 하는 새로운 방법론을 제시하였다. 랜덤포레스트 및 다층퍼셉트론 모델을 활용하여 계통신호를 예측한 후 예측값을 이용하여 새로운 밸브유량계수, 압력손실계수를 산출하여 그 결과를 최초 데이터와 비교하는 평가방법을 도출

하였다.

- 2) 머신러닝 모델을 활용한 유량의 예측 정확도는 가장 높았으며, 밸브개폐율에 대한 예측성능도 높다고 판단된다. 밸브 전·후단 압력차에 대한 예측성능은 상대적으로 낮았으나, 다층퍼셉트론 방법론을 적용하여 비선형 관계를 반영하면서 향상된 결과를 도출하였다. 따라서 실험결과를 바탕으로 밸브개폐율과 유량의 예측에는 랜덤포레스트 모델, 밸브 전·후단 압력차에 대한 예측에는 다층퍼셉트론 모델이 최적의 성능을 제공함을 확인하였다.
- 3) 본 연구에서 제안한 머신러닝 모델이 밸브고유계수와 같은 물리적 불변성을 유지하면서도 높은 예측 성능을 달성함을 확인하였다.
- 4) 다층퍼셉트론 모델 최적화 과정 중 K-Fold 교차검증으로 넘어가는 과정에서 결정계수( $R^2$ )가 소폭 감소하는 경우는 통계적으로 충분히 발생할 수 있는 현상으로, 최적화 전 모델의 성능이 과대 평가되었음을 바로 잡는 과정으로 해석될 수 있고, 이후 하이퍼파라미터 튜닝결과 다시 성능이 좋아지기 때문에 과적합을 조절하면서 최적화된 모델로 수렴해 가는 정상적인 과정이라고 판단할 수 있다.
- 5) 현재 연구에서는 밸브 전·후단 압력차에 대한 예측성능이 밸브개폐율, 유량보다 상대적으로 낮았으며, 이를 해결하기 위해 Transformer 또는 LSTM과 같은 딥러닝 기반 시계열 모델 적용 가능성을 검토할 필요가 있다.

## References

- (1) Kim, J., Kim, D., and Lee, S., 2019, "A Numerical Study for Performance Characteristics of Active Components Used in Nuclear Power Plants", *Journal of Computational Fluids Engineering*, Vol.24, No.1, pp.72-78.
- (2) Kalsi Engineering Inc., 2009, "The AOV/MOV Software with Technology KVAP 2.2 User's Manual"
- (3) Kim, J., Cho, W., Kim, D., Lee, T., Ryu, K., Kim, Y. and Lee, S., 2017, "A Numerical Study on Flow and Thrust Characteristics for Balanced Globe Valve Used in Nuclear Power Plant," *Journal of Computational Fluids Engineering*, Vol.21, No.4, pp.81-88.
- (4) Ko, S. B., Jang, I. H., and Song, S. M., 2012, "Numerical analysis and optimal design to minimize cavitation of vehicle brake valve system," *The KSME Conference*, The Korean Society of Mechanical Engineers, pp. 1248-1250.
- (5) Park, Y. C., Choi, J. S., and Kang, J., 2005, "A optimization of butterfly valve using the characteristic function," *Journal of the OET*, Vol. 19, No. 3, pp. 59-65.
- (6) M. M. Mukaka, 2012, "A guide to appropriate use of Correlation coefficient in medical research", *Malawi Medical Journal*, Vol.24, pp69-71.
- (7) Kim, S., Ahn, H. S., Yoon, C., Chung, J. P., and Nam, S. S., 2024, "Development of Outlier Detection Method Using Statistical Analysis and Correlation Coefficient for Environmental Sensor Data", *Journal of Advanced Navigation Technology*, Vol 28, No. 6, pp.986-994.
- (8) Tukey, J. W., 1977, "Exploratory Data Analysis", Addison-Wesley Publishing Company.
- (9) L. Breiman, 2001, "Random Forests", *Machine Learning*, vol. 45, pp. 5-32.
- (10) Kang, M., 2023, "Renewable Energy Generation Prediction Model using Meteorological Big Data", *Journal of the KIECS*, vol. 18, pp. 39-44.
- (11) I. Oh, 2023, "MACHINE LEARNING"(7st de.), HANBIT Academy, Inc. Published.
- (12) Kim, C., Choi, M., Joo, C., Lee, A., Gun, Y., Cho, S., and Kim, J., 2024, "Data-driven Modeling for Valve Size and Type Prediction Using Machine Learning", *Korean Chem. Eng. Res.*, Vol. 62, No. 3, pp. 214-224.
- (13) ASME/ISA-75.01.01, 2002, "Control Valve Sizing Equations"
- (14) ISO 5167, "Measurement of fluid flow by means of pressure differential devices"
- (15) Lee, Y., Ko, J., and Kim, H., 2017, "Flow Analysis and Profile Modeling for Machine Learning and Performance Evaluation of Control Valve", *Journal of the Comput. Fluids Eng.* Vol.22, No.4, pp. 36-41.
- (16) Kim, J., Park, T., Lee, S., Ryu, H., Lim, T., and Lee, J., 2020, "A Numerical Study on Modeling and Functional Qualification Using Interpretation of Static Diagnosis Test for a Balance Globe Valve in Nuclear Power Plants", *Journal of the Comput. Fluids Eng.* Vol.25, No.3, pp. 14-22.
- (17) Choi, J., Park, S., and Lee, K., 2016, "Flow Analysis and Experimental Study of Globe Valve for Precision Control", *Journal of Korea Academia-Industrial Cooperation Society.*, Vol. 17, No. 7, pp. 734-739.
- (18) Chun, J., Lee, H., Im, S., Kim, D., and Baek, S., 2021, "Comparative assessment of frost event prediction models using logistic regression, random forest, and LSTM networks", *Journal of Korea Water Resour. Assoc.*, Vol. 54, No.9, pp. 667-680.
- (19) Choi, S., Lee, S., and Nam, C., 2024, "Enhanced Machine Learning Preprocessing Techniques for Optimization of Semiconductor Process Data in Smart Factories", *The Journal of The Institute of Internet, Broadcasting and Communication*, Vol.24, No.4, pp. 57-64.