

데이터 마이닝을 활용한 장기저장탄약 상태 결정요인 분석 연구

노유찬* · 조남욱**† · 이동녕***

* 국방기술품질원

** 서울과학기술대학교 산업공학과

*** 인하대학교 항공우주공학과

A Study on Determinants of Stockpile Ammunition using Data Mining

Roh, Yu Chan* · Cho, Nam-Wook**† · Lee, Dongnyok***

* Defense Agency for Technology and Quality

** Dept. of Industrial Engineering, Seoul National University of Science and Technology

*** Department of aerospace engineering, Inha University, Incheon, Republic of Korea

ABSTRACT

Purpose: The purpose of this study is to analyze the factors that affect ammunition performance by applying data mining techniques to the Ammunition Stockpile Reliability Program (ASRP) data of the 155mm propelling charge.

Methods: The ASRP data from 1999 to 2017 have been utilized. Logistic regression and decision tree analysis were used to investigate the factors that affect performance of ammunition. The performance evaluation of each model was conducted through comparison with an artificial neural networks(ANN) model.

Results: The results of this study are as follows; logistic regression and the decision tree analysis showed that major defect rate of visual inspection is the most significant factor. Also, muzzle velocity by base charge and muzzle velocity by increment charge are also among the significant factors affecting the performance of 155mm propelling charge. To validate the logistic regression and decision tree models, their classification accuracies have been compared with the results of an ANN model. The results indicate that the logistic regression and decision tree models show sufficient performance which conforms the validity of the models.

Conclusion: The main contribution of this paper is that, to our best knowledge, it is the first attempt at identifying the significant factors of ASPR data by using data mining techniques. The approaches suggested in the paper could also be extended to other types ammunition data.

● Received 5 February 2020, 1st revised 23 April 2020, accepted 7 May 2020

† Corresponding Author(nwcho@seoultech.ac.kr)

© 2020, Korean Society for Quality Management

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-Commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Key Words: Ammunition Stockpile Reliability Program (ASRP); Data Mining; Logistic Regression; Decision Tree; Artificial Neural Networks(ANN)

1. 서 론

탄약은 지속적으로 사용되지 못하고 단 한 번 기능을 발휘하고 파괴되는 1회성 품목이다. 탄약은 목적에 적합한 환경을 조성하여 사용되기 전까지 성능이나 안전성을 확인하기 어려운 반면, 요구되는 신뢰도는 매우 높다. 기동장비 등의 무기체계는 저장 기간이 거의 존재하지 않고 지속적으로 사용되어 고장 확인이 가능한 반면, 탄약은 저장 기간 내 작동되지 않으므로 저장 중에는 고장을 발견하기가 어렵다(Yoon and Lee, 2012). 그러나 탄약은 시간 경과에 따라 물리적, 화학적 열화 반응으로 성능이 저하되거나 안전성이 떨어지면서 저장수명 동안 신뢰도가 서서히 저하되므로 주기적인 상태 확인 및 점검이 필요하다. 이에 따라 국방품질기술원에서는 저장 탄약 신뢰성 평가(Ammunition Stockpile Reliability Program, 이하 ASRP)를 통해 저장 중인 탄약에 대해 주기적으로 성능과 안전성을 확인하는 신뢰성 평가를 실시하고 있다(Kim, et al., 2010; Yoon and Park, 2014).

ASRP에서는 저장 중인 탄약에 대하여 비기능 시험(육안 검사), 기능 시험(사격 시험) 및 저장분석 시험(이화학 시험)을 실시하여 사용 가능성과 안전성을 평가한다. 평가 결과에 따라 해당 탄약 로트(LOT)에 대해서 계속 저장, 조건부 불출, 우선 불출, 폐기 등으로 탄약 상태 기호(condition code)를 결정한다. ASRP 결과는 저장 탄약의 획득, 저장, 폐기 및 효율적 운용의 지침으로 활용되며, 군 전투력 향상과 국방예산 절감에 기여하고 있다. ASRP를 분석하는 기존 연구로는 탄약 품목별 신뢰도 추정 연구, 추진 장약의 신뢰도와 저장수명 추정 연구, 가속 열화 시험을 통한 신관 구성품의 저장수명 예측 연구, 추진제 수명연장, 의약품 사용 기간 평가 기법을 활용한 저장 탄약의 최초시험연수를 결정하는 연구 등이 있다. 하지만 기존 연구는 대부분 탄약의 신뢰도 및 저장수명 분석에 대한 연구로서 수명에 영향을 미치는 다양한 변수를 분석하기 보다는 저장 기간과 제조년도 등의 요인을 제한적으로 고려함으로써 신뢰도 및 수명 추정에 영향을 미칠 수 있는 다른 유의한 요인을 도출하는 것에 한계가 있었다 (Roh, et al., 2018).

본 연구에서는 ASRP 분석에 데이터 마이닝 기법을 활용하여 탄약 상태 등급에 유의미한 영향을 미칠 수 있는 요인을 도출하고 ASRP 탄약 등급평가 모델을 제시하고자 한다. 이를 통해 ASRP 등급 평가 심의 과정의 가이드라인을 제공함으로써 저장 탄약의 사용, 유지 보관 및 운용 등에 활용하고자 한다. 본 연구에서는 ASRP 대상 탄약 중 1999년부터 2017년까지 239개의 155MM 백색추진장약의 ASRP 데이터를 활용하여 로지스틱 회귀분석, 의사결정나무 분석, 신경망 분석을 실시하고 각 모델의 성능 평가를 실시하였다. 본 연구는 ASRP 분석에 데이터 마이닝을 적용하여 추진장약의 등급 결정에 영향을 미치는 다양한 요인을 도출하였으며 예측정확도가 높은 판별모델을 제시하였다는 것에 기존연구와 차별성이 있다. 본 연구의 구성은 다음과 같다. 2장에서는 선행연구 및 기존 연구와의 차이점을 서술하였다. 3장에서는 연구 방법과 연구모델에 대해 서술하였다. 4장에서는 연구의 내용 및 결과를 서술하고, 5장에서는 결론을 서술하였다.

2. 선행연구

장기저장탄약의 신뢰도와 수명을 예측하는 다양한 연구가 ASRP 시험결과를 활용하여 이루어져 왔으나 기존 연구의 대부분은 시계열 기반의 통계모형을 이용한 연구가 주를 이루고 있다. Lee, et al.(2011)은 155MM 추진장약의

ASRP 결과를 활용하여 추진제의 안정제한량 변화에 따른 시계열분석 방법론을 적용하여 ASRP 저장 탄약 수명을 예측하였으나, 안정제한량과 탄약수명과의 관계만을 분석했다는 한계가 있다. Lee and Hong(2012)은 ASRP 데이터의 효과적 활용을 위해 통계적 기법을 적용한 저장탄약의 시험주기 설정, 시계열분석을 통한 저장탄약 수명 예측, ASRP의 전력효과 분석 등 3가지 방법을 제시하였다. Yoon and Lee(2012)는 접근신관의 ASRP 결과를 활용하여 탄약의 로트 단위별 평가에서 품목 단위 평가 체계로 전환하기 위한 개념을 제안하였다. 이에 필요한 통계적 분석기법을 제안하고, 접근신관의 ASRP 결과를 분석하여 품목 신뢰도를 산출하였다. 또한, 품목별 신뢰도 변화를 예측할 수 있는 평가 방법과 향후 품목별 신뢰도평가를 위한 방안을 고찰하였다. Yoon(2013)은 ASRP 155MM 추진장약 M4계열에 대해서 13년도에 추가적으로 시험 평가한 시험자료를 활용하여 품목 단위로 신뢰도를 산출하였다. 이 연구는 군 저장환경, 제조 시 품질 수준 등의 조건을 고려하지 않았으므로 신뢰성 있는 정확한 수명 예측에는 한계가 있지만, 탄약의 수명 추정에 활용 가능한 기법을 검토하고, 검토된 기법에 따라 저장수명을 추정하고자 하였다. Yoon and Park(2014)은 155MM 추진장약 KM4A2 ASRP 결과로부터 품목 단위 신뢰도 및 저장수명을 산출하는 분석기법을 연구하여, 추진장약 데이터를 계량치와 계수치 자료로 구분하여 제시하였고, 이를 통해 신뢰도 및 저장수명을 산출하였다. Yoon(2014)은 2000년부터 2013년도 중에 실시한 ASRP 업무를 통해 얻어진 자료와 WRSA 인수 탄약 평가에서 얻은 자료를 활용하여 155MM 추진장약 M119계열의 품목 단위 저장 신뢰도와 저장수명을 산출하였다. 군 저장환경, 제조 시 품질 수준 등의 조건을 고려하지 않았으므로 신뢰성 있는 정확한 수명 예측에는 한계가 있지만, 탄약의 수명 추정에 활용 가능한 기법을 검토하고, 검토된 기법에 따라 저장수명을 추정하고자 하였다. Choi, et al.(2014)는 단기추진제로 구성된 국내 105MM 고폭탄의 ASRP 자료를 활용하여 탄약의 저장수명과 관련된 속성을 찾고, 분석 결과를 통해 단기추진제의 저장수명을 예측하여 신뢰성 확보에 기여하고자 하였다. 이를 위해 잔류 안정제한량 속성을 종속변수로 두고, 제조년도 및 저장 기간 속성을 독립변수로 활용하여 각각 회귀분석 및 신뢰도 분석한 결과 유의성을 확인하였다. Lee, et al.(2015)는 단기추진제를 사용하여, 1단계에서는 고온 가속 노화 시험법을 적용하여 저장온도별 안정제한량 감소 데이터를 구하고, 2단계에서는 고온 가속 노화 시험을 통해 얻어진 저장온도별 안정제한량 감소 데이터를 이용하여 n차 반응속도 식으로 모델링하고, 최적의 반응차수를 구하였다. 이를 통해 0차와 1차의 반응차수와 신뢰수준을 비교하여 단기추진제의 저장수명 예측 신뢰성 확보에 기여하고자 하였다. Jeong and Lee(2017)는 10년 이상 저장된 유도탄의 야전 점검 데이터를 수집하여 실제 저장 신뢰도를 분석하고 현재 실시 중인 검사주기의 타당성 검토와 실질적인 저장 신뢰도 확인방법을 제시하고자 하였다. 연구에서 저장 신뢰도를 분석한 결과, 검사주기는 18년 이상으로 연장 가능성이 확인되었고, 이에 따른 검사주기의 재설정 및 실질적인 저장 신뢰도 확인방법을 제안하였다.

살펴본 바와 같이 기존 선행연구에서 활용된 방법론은 시계열분석, 회귀분석 등을 활용하여 안정제한량의 감소 추세를 통해 간접수명을 추정하거나 분포를 가정하여 불량수의 증가 추세를 이용하는 방법을 통해 수명을 추정하는 방법이 주를 이루었다. 안정제한량은 시간이 지남에 따라 감소하는데, 0.3 이하가 되면 매우 불안정한 상태로 개수정비 혹은 폐기하게 된다. 시간에 따른 안정제한량의 감소가 일정한 양상을 나타내기 때문에, 안정제한량의 연간 감소량을 회귀계수로 추정한다. 회귀분석 결과를 기반으로 안정제한량의 한계지점인 0.3 이하가 되는 저장년도를 탄의 수명으로 추정하였다. 하지만 이러한 방법론은 직접수명이 아닌 안정제한량을 간접수명으로 활용하였다는 한계가 있다. 고장 수를 출력변수로 활용한 대부분의 기존 연구는 고장 수의 분포를 가정하고, 저장 기간에 따른 고장 수를 추정하였다. 이러한 연구들은 탄의 성능이 끝나는 시점(고장)까지의 기간인 직접수명을 추정하였으며, 다양한 요인을 고려할 수 없다는 한계가 있다. 따라서 탄의 수명에 영향을 미칠 수 있는 요인을 다각도로 고려하여 판별하고 이를 반영할 수 있는 높은 예측력을 가진 모델의 개발이 필요하다.

3. 연구방법론

3.1 연구절차

본 연구는 1) 데이터 수집과 데이터 정제, 2) 데이터 마이닝 모델의 개발, 3) 모델의 검증 및 비교의 순서로 진행되었다. 본 연구에서 사용된 데이터는 1999년부터 2017년까지의 ASRP를 통해 얻은 총 277개 데이터 중 결측치를 제외한 총 239개의 155MM 백색추진장약 데이터이다. 비정형 ASRP 시험결과를 분석에 활용하기 위해서 데이터를 수치형 혹은 범주형 변수로 생성하는 전처리 작업을 실시하였다. 우선, 비정형 텍스트 형태로 입력이 되어있는 추진장약의 일반적 특성을 DODIC(Department Of Defence Identification Code), 국내외품, 점화제 종류에 따라 분류하였으며, 각 시험에서의 나타난 결점은 샘플 시료수 대비 결점의 비율로 바꾸었다. 이를 통해 탄약의 일반적 특성, 비기능 시험, 기능 시험, 저장분석 시험 결과에 따라 데이터 처리를 하여 13개의 입력변수를 설정하였다. 출력변수로 설정한 탄약 상태 기호(condition code)는 CC-A 등급 이외에는 추진장약에 하자가 있는 것으로 판단하여, 추진장약의 사용이 양호한 집단(0)과 그렇지 않은 불량 집단(1) 두 가지로 나누었다. Table 1은 데이터 정제를 통해 만든 입력 변수 목록이다.

Table 1. Input Variables

Category	Variable Name	Type	Description	
Input Variables	General characteristic	DODIC	categorical	item classifier
		Origin of production	binary	domestic or foreign production
		Ignition powder type	categorical	types of ignition powder
		Production year	ordinal	year produced
	Visual inspection	Critical defects rate (치명결점 비율)	continuous	fraction of critical defects
		Major defects rate (중결점 비율)	continuous	fraction of major defects
		Minor defects rate (경결점 비율)	continuous	fraction of minor defects
	Functional test	Major defects rate (중결점 비율)	continuous	fraction of major defects
		Muzzle velocity by base charge(저장약 탄속)	continuous	average muzzle velocity by base propelling charge
		Muzzle velocity standard deviation by base charge (저장약_탄속표준편차)	continuous	standard deviation of muzzle velocity by base propelling charge
		Muzzle velocity by increment charge(고장약 탄속)	continuous	average muzzle velocity by increment propelling charge
		Muzzle velocity standard deviation by increment charge(고장약_탄속표준편차)	continuous	standard deviation of muzzle velocity by increment propelling charge
	Stockpile Analysis Test	Content of stabilizer (안정제 함량)	continuous	content of stabilizer in propelling charge

3.2 연구모델

본 연구에서는 데이터 마이닝 기법인 로지스틱 회귀분석, 의사결정나무 분석, 신경망 분석을 통하여 탄약 상태에 영향을 미치는 요인을 찾고 각 모델의 성능 비교를 진행하였다. 본 논문에서 적용된 모델은 다음과 같다.

3.2.1 로지스틱 회귀분석

로지스틱 회귀분석은 개별 데이터가 어느 집단으로 분류될 수 있는가 분석하여 사건 발생확률(1)과 발생하지 않을 확률(0)을 분석할 수 있다. 본 연구에서는 출력변수인 추진장약의 양호집단과 불량집단으로 이항(binary) 구성된 로짓모형을 사용하였다. 로지스틱 회귀분석의 적합성은 Hosmer& Lemeshow 검정 등을 통해 검증하였다.

3.2.2 의사결정나무 분석

의사결정나무 분석은 데이터를 가장 잘 분류할 수 있는 분리기준을 찾아 분리 마디를 생성하여 데이터를 분류하는 방법이다 (Choi and Riew, 2016). 의사결정나무 분석은 주요 요인변수를 찾아내는 데 유용하다 (Choi et al. 2015). 본 연구에서는 의사결정 나무 분석을 위해 이진(binary) 분류의 알고리즘인 CART(Classification and Regression Trees; CART) 방법을 사용하였다. 그리고 CART 알고리즘의 평가 지수로는 노드의 다양성을 줄이는 방법으로 대상 변수의 값에 따라 하위노드의 동질성이 최대가 되도록 분할하는 Gini index를 이용하였다. 또한, 트리 깊이는 4, 터미널 노드 수 5, 상위노드의 사례 수는 8, 하위노드 사례 수는 4로 설정하였다.

3.2.3 신경망 분석

본 연구는 신경망 분석을 위해 입력변수의 값을 기준으로 출력변수의 예측 오차를 최소화하는 예측모형을 생성하는 다층퍼셉트론(Multi Layer Perceptron ;MLP) 네트워크 모형을 사용하였다. 그리고 신경망 내의 연결이 피드백 루프없이 입력층에서 출력층으로 전방향으로 흐르는 전방향 구조를 선택하였고, 각 층에 설정되는 활성화 함수는 쌍곡탄젠트 함수를 사용하였다. 쌍곡탄젠트 함수는 시그모이드 함수의 크기와 위치를 조절(rescale and shift)한 함수로 쌍곡탄젠트 함수는 시그모이드를 활성화함수로 썼을 때보다 학습 수렴 속도가 빠른 장점이 있다. 모델의 훈련은 모든 훈련데이터 레코드를 전달한 후에만 시냅스 가중치를 업데이트하는 배치 훈련방법을 이용하였다. 그리고 신경망 분석의 적합성은 오차 제곱의 합의 계산을 통해 실시하였다. 또한, 본 연구에서는 1개의 입력층, 1개의 은닉층, 1개의 출력층으로 구성된 3계층 분석모형을 사용하였다. Jung and Jung(2009)은 층과 노드가 많아질수록 복잡한 과정을 더욱 정확하게 처리할 수 있으나, 너무 많은 층과 노드를 사용하면 과학습(overtrained)으로 인한 과적합(overfitting) 상태를 초래할 수 있고, 반대로 너무 적은 층과 노드를 사용하면 불충분합(underfitting)을 초래할 수 있다 하였다. 또한, 데이터와 원하는 결과값에 따라 최적의 은닉 노드의 수는 차이를 가지므로, 상황에 적합한 최적의 은닉 노드의 수는 설정할 수 있지만, 모든 상황에 적용되는 최선의 신경망은 존재하기 어렵다고 하였다(Jung and Jung, 2009). 따라서 본 연구에서는 13개의 입력변수를 입력층에 입력하고 은닉층의 노드 수는 변경해가면서(5개, 10개, 15개, 20개, 25개, 30개, 35개) 5회 반복하여 실시하여 분류율이 높은 노드 수를 찾았다.

3.2.4 모델 비교

로지스틱 회귀분석, 신경망 분석, 의사결정나무 분석을 각각 실시하고 모델의 검정력은 정오분류표(confusion matrix)와 수신자 판단 곡선(Receiver Operating Characteristic curve; ROC 곡선)을 이용한 ROC 곡선 아래 면적

(Area Under the ROC Curve; AUC)를 확인하였다. 데이터는 훈련데이터 160개(66.9%), 검정데이터 79개(33.1%)로 구분하여 분석하였다.

4. 연구결과

4.1 로지스틱 회귀 분석

로지스틱 회귀분석을 통해 어떤 입력변수들이 추진장약의 양호집단(0)과 불량집단(1) 분류에 영향을 미치는지 분석하였다. 분석을 위해 총 239개의 데이터를 활용하였으며, 이를 훈련데이터 160개(66.9%)와 검정데이터 79개(33.1%)로 임의적으로 다시 나누어 분석함으로써 로지스틱 회귀분석의 분류능력을 검증하였다. 239개의 데이터 중에서 양호한 집단의 개수는 111개이고, 불량 집단의 개수는 128개수로 나타났다. 대상 탄약은 평균 42년 동안 저장되었던 관계로 불량 개수가 상대적으로 많이 발생했기 때문에 데이터 불균형 문제는 발생하지 않았다.

분석 결과는 Table 2에 제시되었다. $-2 \text{ Log 우도}(-2 \log \text{ likelihood}; -2LL)$ 는 모형의 적합도를 나타내는데, $-2LL$ 이 낮을수록 모형적합도가 높다고 할 수 있다. 본 연구에서는 입력변수를 포함하지 않고 상수항만으로 구성된

Table 2. Results of Logistic Regression

variables	description	B	S.E.	Wald	DoF	sig.	Exp(B)
production year	year when an item has been produced	-.036	.059	0.365	1	.546	.965
ignition powder type	types of ignition powder	2.024	.980	4.263	1	.039**	7.566
major defects rate (%) of visual test	fraction of major defects	.507	.123	16.876	1	.000***	1.660
major defects rate(%) of functional test	fraction of major defects	.415	.214	3.771	1	.052*	1.514
muzzle velocity standard deviation by base charge	standard deviation of muzzle velocity by base propelling charge	3.02	.730	17.149	1	.000***	20.556
muzzle velocity by increment charge	average muzzle velocity by increment propelling charge	-.268	.125	4.614	1	.032**	.765
muzzle velocity standard deviation by increment charge	standard deviation of muzzle velocity by increment propelling charge	2.31	.695	11.040	1	.001***	10.077
constant term		218.12	113.81	3.673	1	.055*	5.34E94

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

$-2 \log \text{ likelihood}(-2LL) = 75.720$, Cox & Snell $R^2 = 0.595$, Nagelkerke $R^2 = 0.796$

Hosmer & Lemeshow test(검정) $\chi^2 = 6.050$ (d.f.=8, $p = 0.642$)

모형의 $-2LL$ 이 220.204였으나 입력변수 모두를 포함한 모형의 경우, $-2LL$ 이 75.720으로 낮아져 입력변수들이 추진장약의 상태 결정에 영향을 미치는 것을 확인할 수 있었다. 또한, Hosmer와 Lemeshow 검정의 경우 유의확률이 0.05보다 크면 모형이 적합하다고 할 수 있다. 따라서, $chi^2=6.050$ (d.f.=8, $p=0.642$)이므로 모형의 적합도는 수용할 만한 수준이라 할 수 있다. 그리고 입력변수들이 출력변수를 설명하는 정도를 나타내는 R^2 (결정계수)의 값은 Cox&Snell의 R^2 값이 0.595, Nagelkerke의 R^2 이 0.796으로 나왔다. 이는 로지스틱 회귀분석에서 사용된 입력변수들에 의해 출력변수 분산의 59.5~79.6% 정도가 모형에 의해 설명된다고 할 수 있음을 보여준다.

출력변수의 영향을 미치는 입력변수의 통계적 유의성을 검증하는 값인 Wald 통계량과 유의확률을 살펴보면 다음과 같다. 점화제종류(ignition power type), 비기능시험 중결점 비율(major defect rate of visual inspection), 저장약 탄속 표준편차(muzzle velocity standard deviation by base charge), 기능시험 고장약 탄속(muzzle velocity by increment charge), 기능시험 고장약 탄속 표준편차(muzzle velocity standard deviation by increment charge)가 5% 유의수준에서 추진장약의 상태에 유의한 영향을 미치는 것으로 분석되었다. 기능 시험 중결점 비율(major defect rate of function test)은 5% 유의수준에는 근소하게 미치지 못하는 못하였으나 유의한 변수로 판단되며 제조년도는 유의하지 않은 변수로 판단된다.

4.2 의사결정나무 분석

의사결정나무 분석을 통해서 추진 장약이 양호집단(0)과 불량집단(1)으로 분류되는 추진장약 상태 결정요인을 알아보고자 하였다. 이를 위해 CART 확장방법에 지니계수를 사용한 이진 분류 형태의 의사결정나무 분석을 실시하였고, 훈련데이터 160개(66.9%)와 검정데이터의 79개(33.1%)를 이용한 결과는 Table 3에 요약하였다. 입력변수의 중요도가 가장 높은 것은 비기능 시험에서의 중결점 비율(major defect rate)이었으며, 기능시험에서의 고장약 탄속(muzzle velocity by increment charge), 저장약 탄속 표준편차(muzzle velocity standard deviation by base charge), 고장약 탄속 표준편차(muzzle velocity standard deviation by increment charge), 저장약 탄속(muzzle velocity by base charge), 비기능시험에서의 치명결점 비율(critical defects rate)의 순으로 나타났다.

Table 3. Results of decision tree

input variables	significance	normalized significance
major defects rate of visual inspection	.149	100.0%
muzzle velocity by increment charge	.089	59.8%
muzzle velocity standard deviation by base charge	.081	54.1%
muzzle velocity standard deviation by increment charge	.058	38.8%
muzzle velocity by base charge	.048	32.0%
critical defects rate of visual inspection	.043	29.1%

4.3 신경망 분석

로지스틱 회귀분석과 의사결정나무 분석은 분석결과에 영향을 미치는 변수를 판별할 수 있다는 이점이 있으나 예측정확도가 다소 떨어진다는 단점이 존재한다. 본 논문에서는 로지스틱 회귀분석과 의사결정나무 분석을 통해 얻은

모델의 성능 검증 및 정확도 비교를 위해 신경망 분석을 실시하였다. 총 239개의 데이터를 활용하였으며, 이를 훈련 데이터 160개(66.9%)와 검증데이터 79개(33.1%)로 나누어 분석함으로써 신경망 분석의 분류능력을 검증하였다. 일반적으로 신경망은 학습된 내용을 다루는 것에 있어 인간이 쉽게 이해하기 어려운 면이 많은 것과 과학습에 따른 과적합이 발생하기 쉽다는 문제점을 가지고 있다. 과적합은 은닉 노드의 수에 민감한 때문에, 본 연구에서는 신경망 분석은 1개의 은닉층에서 노드 수를 변경해가면서 총 5회 반복 시험을 하여 Table 4에 결과를 제시하였다. 신경망은 입력층, 은닉층, 그리고 출력층으로 구성되어 다양한 모형을 포함하는 매우 유연한 모델이지만 결과 해석은 쉽지 않다. 로지스틱 회귀분석과 같이 계수들에 대한 간편한 해석이 불가능하여 어떤 입력변수가 중요한지, 그것들이 어떻게 상호 작용하는지 제시하기가 어렵다. 따라서 연구자는 예측과 해석적인 측면을 잘 고려하여 최종적인 모형을 선택하는 것이 바람직하다(Kang, et al., 2014).

Table 4. Results of ANN

No.	# of hidden layers	Training data		Test data		AUC
		Error Sum of Squares	Classification accuracy(%)	Error Sum of Squares	Classification accuracy(%)	
1	5	7.9818	95.90	4.9276	94.68	0.9928
2	10	8.4018	96.26	5.3744	95.16	0.9934
3	15	6.9112	98.40	5.9496	93.18	0.9944
4	20	7.9806	97.78	5.9328	94.68	0.9962
5	25	7.2778	98.28	6.2074	94.18	0.9946
6	30	6.7200	98.52	6.6162	92.92	0.9950
7	35	6.3744	98.50	6.5528	92.40	0.9946

4.4 각 모델의 분석 결과 비교

본 연구에서는 로지스틱 회귀분석, 의사결정나무 분석, 신경망 분석의 결과 비교를 통해 추진장약의 상태 결정에 영향을 미치는 요인을 찾고, 각 모형의 성능을 비교해 보고자 하였다. 이를 위해 훈련데이터 160개(66.9%)와 검증데이터 79개(33.1%)를 임의로 선택한 후 각각의 모형에 대입하였다.

Table 5는 로지스틱 회귀분석과 의사결정나무 분석을 통해 얻은 모델의 성능과 검증을 위해 신경망 분석과 비교한 결과이다. 신경망 분석의 분류정확도(accuracy)는 훈련데이터에서 96.3%, 검증데이터에서 97.4%로 가장 높은 것으로 나타났으며, F1-score도 가장 높은 것으로 나타나 분류정확도 측면에서는 신경망 분석이 가장 우수하였다. 로지스틱 회귀분석과 의사결정나무 분석의 경우 신경망 분석의 분류성능에 근사한 성능을 보여주었다. 이는 로지스틱 회귀분석 모형과 의사결정나무 모형의 적합성을 시사하며 앞 절에서 제시한 요인의 유효성을 뒷받침한다. 각 모델의 적합도를 알아보기 위한 ROC 곡선을 이용한 AUC를 확인한 결과에서도 각 분석 모델의 값이 모두 0.9 이상을 보였다.

Table 5. Comparison of Models

Result	Logistic Regression		Decision Tree		ANN(# of hidden layers=20)	
	Train	Test	Train	Test	Train	Test
Accuracy	0.905	0.873	0.900	0.937	0.963	0.974
F1-score	0.897	0.872	0.983	0.938	0.958	0.975
AUC	0.963		0.931		0.994	

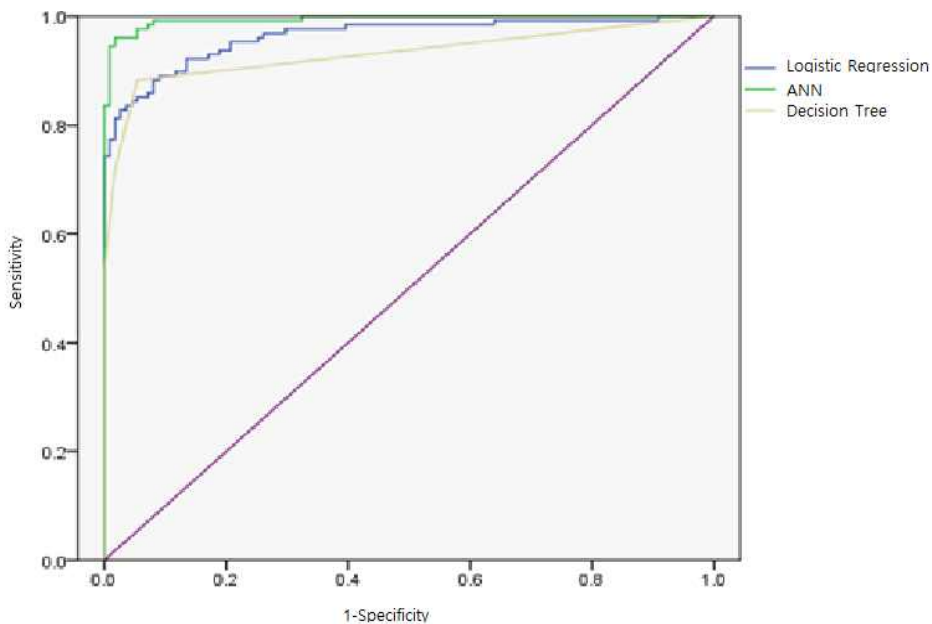


Figure 1. Comparison of AUC

5. 결 론

본 연구는 ASRP 데이터에 데이터 마이닝을 활용하여 추진장약 성능에 영향을 미치는 요인을 찾고, 추진장약의 등급평가 모델을 만들고 활용을 시도하고자 하였다. 이를 통해, ASRP 등급평가 심의 과정 시 추진장약 사용과 관련한 자료 해석에 대한 시간 절약 및 효율성을 제공하고, 군에는 ASRP 대상 탄약의 사용, 유지 및 관리 등의 사항에 있어 참고자료로써 활용 가능성을 제공하고자 하였다. 분석결과, 비기능 시험의 중결점 비율이 로지스틱 회귀분석과 의사결정나무 분석에서 모두 유의한 것으로 나타났으며 특히 의사결정나무 분석에서는 가장 중요한 요인으로 판별되었다. 기능 시험의 저장약 탄속, 고장약 탄속, 고장약 탄속 표준편차도 로지스틱 회귀분석과 의사결정나무 분석에서 모두 유의한 것으로 나타났다. 로지스틱 회귀분석과 의사결정나무 분석 모형의 성능을 비교하기 위해 신경망 분석과 분류성능을 비교실험한 분석에서는 약 90%의 분류정확도와 0.9 이상의 AUC를 보여줌으로써 모델 적합도가 우수한 것을 확인하였다.

본 연구는 실제 다양한 탄약 상태 기호 결과를 적용하지 못하고 추진 장약의 사용이 양호한 집단(0)과 그렇지 않은 불량 집단(1)으로 이진 분류하였다. 다항 로지스틱 회귀분석을 통해 분류의 세분화로 업무의 적용 가능성을 확대하는 것은 추후연구가 될 수 있다. 추진장약 이외의 다른 탄종에서의 분석 가능성 및 수명 예측과 관련하여 데이터 마이닝을 활용한 연구 가능성을 확대해 나감으로써 기존 연구와의 결과 비교 및 다양한 분석을 시도해볼 필요가 있다.

REFERENCES

- Choi, H. D., Riew, M. C. 2016. An Analysis of Ordinary Mail Service Quality Attributes using Kano Model and Decision Tree Model. *Journal of the Korean Society for Quality Management* 44(4):883-895.
- Choi, J. S., Han, J. T., Kim, M.J., Jeong, J. 2015. Developing the High Risk Group Predictive Model for Student Direct Loan Default Using Data Mining. *Journal of the Korean Data and Information Science Society* 26(6):1417-1426.
- Choi, M. J., Park, H. J., Yang, J. K., Baek, J. H. 2014. A Study on the Shelf-Life Prediction of the Domestic Single Base Propellants Ammunition: Based on 105mm High Explosive Propellants. *Journal of the Society of Korea Industrial and Systems Engineering* 37(3):36-42.
- Jeong, S. H., Lee, S. B. 2017. Analysis of $\Delta\Delta$ Guided Missile Inspection Period Based on Storage Reliability. *Journal of the Korea Academia-Industrial Cooperation Society* 18(4):592-598.
- Jung, S. H., Jung, Y. G. 2009. Neural Networks for Optimal Hidden Nodes Decision. *Proc. of the 36th KIISE Fall Conference* 36(2C):334-337.
- Kang, H. C., Kim, E. S., Lee, S. G., Lee, S. J., Chae, Y. S., Choi, Y. S., Choi, B. J., Choi, J. H., Han, S. T. 2014. *Functionality and Usage of SAS Enterprise Miner for Big Data Analysis*. Paju: Freeacademy
- Kim, Y. S., Choi, J. S., Kim, D. H., Noh, J. Y., Kim, K. Y., Lee, I. H., Yu, J. M., Kim, Y. H., Jeong, U. G., Kim, W. S., Gu, -W. S., Kim, S. J., Lee, D. C., Yoon, K. S., Kim, J.H. 2010. *Introduction to Defense Quality Management*. Hyungseul Publishing.
- Lee, J. W., Hong, Y. K. 2012. A Study on the Effective Management of Artillery Ammunition Using ASRP Data -The Case of Test Interval Determination, Shelf-life Prediction, Force Effectiveness Analysis-. *Journal of the Korea Academia-Industrial Cooperation Society* 13(9):4349-4358.
- Lee, J. W., Kim, H. B., Kim, Y. I., Hong, Y. G. 2011. Prediction of the Shelf-life of Ammunition by Time Series Analysis. *Journal of the Military Operations Research Society of Korea* 37(1):39-47.
- Lee, S. B., Seo, J. W., Choi, K. S., Kim, S. B. 2015. The Shelf-life Prediction of Single-Base Propellants by Applying the Kinetic Model of n-th Order. *Journal of the Korea Academia-Industrial Cooperation Society* 16(5): 3633-3642.
- Noh, Y. C., Lee, D. N., Cho, N. W. 2018. A Study on Determinants of Stockpile Ammunition Using Data Mining. *The Korean Society for Quality Management. Proceedings*. p. 20.
- Yoon, K, S. 2013. The Reliability and Shelf-life Estimation on 155mm Propelling Charge M4 Series Defense Agency for Technology and Quality. *Research Report DTAQ-13-3774-R*.
- Yoon, K, S. 2014. The Estimation of Stockpile Reliability and Shelf-life for Charge Propelling 155 Millimeter: M119 Series Defense Agency for Technology and Quality, *Research Report DTAQ-14-3981-R*.
- Yoon, K. S., Lee, J. C. 2012. A Case Study on the Reliability Assessment of Stockpile Ammunition. *Journal of the Korean Society for Quality Management* 40(3):259-269.

Yoon, K. S., Park, S. W. 2014. A Study on the Estimation of Shelf-Life for 155mm Propellating Charge KM4A2 Using ASRP's Data. Journal of the Korean Society for Quality Management 42(3):291-300.

저자소개

노유찬 서울과학기술대학교에서 기계공학학사, 산업정보시스템 석사학위를 취득하였다. 현재는 국방기술품질원 기술 정보확산팀에서 연구원으로 재직 중이다. 주요 관심사는 ASRP 및 통계기법이다.

조남욱 서울대학교 산업공학과를 졸업하고 동대학원에서 석사학위를 취득 후 Purdue 대학교에서 산업공학 박사학위를 취득하였다. Lucent Technologies에서 Member of Technical Staff로, 삼성SDS에서 책임 컨설턴트로 근무하였다. 현재 서울과학기술대학교 산업공학과 교수로 재직 중이다. 주요 관심분야는 비즈니스 프로세스 관리와 사회연결망 분석이다.

이동녀 인하대학교 항공우주공학 학사를 취득하였고 현재 국방기술품질원에서 연구원으로 재직하며 인하대학교 항공 우주공학 석박사 통합과정 중에 있다. 주요 관심사는 저장탄약신뢰성평가와 점탄성재료에 대한 구조해석이다.