

# 불균형데이터의 비용민감학습을 통한 국방분야 이미지 분류 성능 향상에 관한 연구

정미애<sup>1)</sup> · 마정목<sup>\*2)</sup>

<sup>1)</sup> 울곡이이함

<sup>2)</sup> 국방대학교 국방과학학과

## A Study on the Improvement of Image Classification Performance in the Defense Field through Cost-Sensitive Learning of Imbalanced Data

Miae Jeong<sup>1)</sup> · Jungmok Ma<sup>\*2)</sup>

<sup>1)</sup> *Yulgok Yii, Republic of Korea Ship*

<sup>2)</sup> *Department of Defense Science, National Defense University, Korea*

(Received 17 September 2020 / Revised 21 January 2021 / Accepted 7 May 2021)

### Abstract

With the development of deep learning technology, researchers and technicians keep attempting to apply deep learning in various industrial and academic fields, including the defense. Most of these attempts assume that the data are balanced. In reality, since lots of the data are imbalanced, the classifier is not properly built and the model's performance can be low. Therefore, this study proposes cost-sensitive learning as a solution to the imbalance data problem of image classification in the defense field. In the proposed model, cost-sensitive learning is a method of giving a high weight on the cost function of a minority class. The results of cost-sensitive based model shows the test F1-score is higher when cost-sensitive learning is applied than general learning's through 160 experiments using submarine/non-submarine dataset and warship/non-warship dataset. Furthermore, statistical tests are conducted and the results are shown significantly.

Key Words : Imbalanced Data(불균형데이터), Cost-sensitive Learning(비용민감학습), CNN(컨볼루션 신경망)

### 1. 서론

4차 산업혁명의 중심기술 중 하나인 딥러닝 기술이

학계 및 각 산업분야에 적용되면서, 국방분야 또한 딥러닝 기술을 적용하고자 하는 시도가 계속되고 있다. 특히, 딥러닝 중에서도 이미지 인식분야는 국방 경계 감시 시스템<sup>[1]</sup>, 군함/비군함 분류<sup>[2]</sup>, 군함 함종 식별 시스템<sup>[3]</sup>, 잠수함 비음향 탐지를 위한 마스트 식별<sup>[4]</sup> 등에 다양하게 연구되면서 딥러닝 기술의 국방분야 적

\* Corresponding author, E-mail: jxm1023@gmail.com  
Copyright © The Korea Institute of Military Science and Technology

용가능성에 대한 기대를 한층 높이고 있다.

이미지 인식은 입력 데이터가 주어졌을 때 해당 데이터의 클래스를 예측하는 분류 문제이다. 예측을 할 수 있는 모델을 구축하기 위해서는 훈련데이터로 모델을 학습시켜야 하며, 이상적으로는 훈련데이터의 클래스별 데이터 수가 유사하여야 한다. 클래스별 수의 차이가 발생하는 데이터를 불균형데이터라고 하는데<sup>[5]</sup>, 어떠한 데이터셋이라도 클래스별 데이터 수가 조금씩차이는 날 수 있지만, 머신러닝이나 딥러닝에서의 불균형데이터는 클래스 간 데이터 수의 불균형이 현저하게 나타나는 경우를 의미하며<sup>[6]</sup>, 본 연구에서는 1.5 이상의 불균형 비율을 가진 경우를 불균형데이터로 정의하였다. 불균형데이터에서는 데이터 간의 분류 경계선이 적절하게 형성되지 못하고 데이터가 많은 다수클래스에 편향되어 형성되게 되며, 이 경우 소수 클래스에 대한 예측 성능이 낮아진다는 문제가 있다.

현실문제에서는 불균형데이터인 경우가 다수 존재한다. 예를 들어 제조업에서 불량품을 판정하는 경우를 딥러닝을 통해 학습시킨다면, 정상인 상품이 불량인 상품보다 훨씬 많으므로 불균형데이터에 해당한다. 이 경우 정상인 상품에 편향되어 학습되게 되고 학습된 모델은 불량인 상품에 대한 낮은 식별율을 보이게 된다. 압 진단도 유사한데, 음성인 경우가 양성보다 훨씬 많으므로 음성에 편향되어 학습되게 되고 불균형데이터를 적절하게 처리하지 않으면 모델의 예측 성능은 낮아지게 된다.

국방분야에서 이미지 분류를 위해 딥러닝을 적용하고자 하는 문제에서도 이와 유사하게 불균형데이터를 처리해야 하는 경우가 빈번하게 발생한다. 경계 감시 시스템에서는 비침입 배경의 데이터가 감시하고자 하는 침입자의 데이터에 대비해 다수이다. 잠수함의 비음향 탐지를 위해 잠수함 마스트를 식별하는 문제에서도 잠수함 마스트의 이미지는 다른 유사데이터에 비해 획득하는 것이 어렵기 때문에 불균형한 데이터 세트가 될 수 밖에 없다. 균함과 비균함을 식별하고자 하는 문제도 유사한데, 균함이 항해를 위해 TV카메라 등의 영상장비를 작동 중에 얻을 수 있는 데이터는 균함에 비해 비균함인 여객선, 유조선, LNG선, 어선 등의 데이터가 훨씬 많다.

최초로 모델을 구축할 때 클래스별로 동일한 양의 데이터셋을 인위적으로 구성한 후 학습할 수도 있겠지만, 모델 성능 향상을 위해서는 데이터를 지속적으로 수집하여 업그레이드해야 한다. 경계 감시 시스템의 TOD나 함정의 TV카메라 등을 통해 수집한 데이터를 통해 모델을 업그레이드한다면, 자연스럽게 얻어지는 데이터는 불균형데이터가 되기 쉬울 것이다.

이에 본 연구는 국방분야에는 최초로 이미지 분류의 불균형데이터 문제를 개선하기 위한 방안으로 클래스에 따라 서로 다른 비용을 부여하는 비용민감학습(Cost-sensitive learning) 기반 모델을 제시한다. 국방분야에서의 불균형데이터 문제와 관련한 연구가 미흡한 상황에서, 잠수함 마스트 식별을 위한 잠수함/비잠수함 데이터셋과 균함/비균함 분류를 위한 데이터셋으로 CNN(Convolutional Neural Network)을 이용한 이미지 분류를 시도할 때, 비용민감학습 기반 모델이 일반 모델보다 성능이 우수함을 보였다. 2장에서는 불균형데이터의 문제점과 이를 처리하기 위한 기술, 선행된 유사연구에 대해 설명하였고, 3장에서는 비용민감학습 기반 이미지 분류모델을 구축한 후 여러 데이터셋을 활용하여 일반 모델들의 예측 성능과 비교 실험을 진행하였다. 4장에서는 예측 성능의 차이가 통계적으로 유의미한지 분석하였다. 5장에서는 연구결론 및 제한사항을 제시하였다.

이제 본 연구는 국방분야에는 최초로 이미지 분류의 불균형데이터 문제를 개선하기 위한 방안으로 클래스에 따라 서로 다른 비용을 부여하는 비용민감학습(Cost-sensitive learning) 기반 모델을 제시한다. 국방분야에서의 불균형데이터 문제와 관련한 연구가 미흡한 상황에서, 잠수함 마스트 식별을 위한 잠수함/비잠수함 데이터셋과 균함/비균함 분류를 위한 데이터셋으로 CNN(Convolutional Neural Network)을 이용한 이미지 분류를 시도할 때, 비용민감학습 기반 모델이 일반 모델보다 성능이 우수함을 보였다. 2장에서는 불균형데이터의 문제점과 이를 처리하기 위한 기술, 선행된 유사연구에 대해 설명하였고, 3장에서는 비용민감학습 기반 이미지 분류모델을 구축한 후 여러 데이터셋을 활용하여 일반 모델들의 예측 성능과 비교 실험을 진행하였다. 4장에서는 예측 성능의 차이가 통계적으로 유의미한지 분석하였다. 5장에서는 연구결론 및 제한사항을 제시하였다.

## 2. 문헌연구

### 2.1 불균형데이터의 문제점과 처리기술

머신러닝이나 딥러닝에서의 분류문제는 학습데이터 세트로부터 클래스별로 적절한 분류경계선을 찾고, 이 모델로 새로운 데이터의 클래스를 예측하는 것이다. 클래스 별 특징이 분명하고 데이터의 수가 균등하다면 Fig. 1에서 보는 것과 같이 이상적인 분류경계선

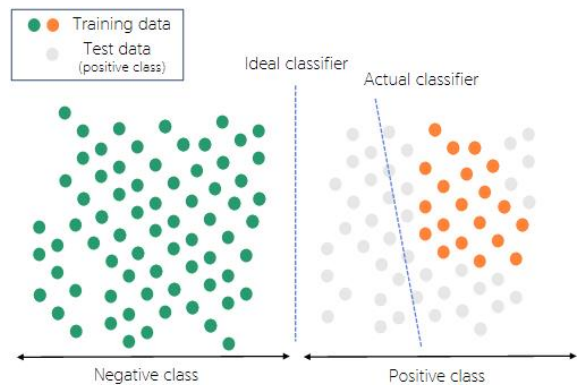


Fig. 1. Classifier in imbalanced data

(Ideal classifier)이 형성되겠지만, 데이터가 하나의 클래스에 편중된다면 다수클래스에 편중된 오른쪽의 분류경계선을 형성하게 된다. 이 경우 형성된 모델에서는 예측하고자 하는 데이터인 회색의 데이터를 음성(Negative class)으로 오분류하게 되는 것이다<sup>[7]</sup>.

불균형데이터로 인해 발생하는 문제를 극복하기 위한 기법은 크게 데이터 수준의 방법과 알고리즘 수준의 방법이 있다<sup>[8]</sup>. 데이터 수준의 방법은 데이터 수의 균형을 맞추어 학습하는 방법으로 샘플링(Sampling) 기법이 이에 해당한다. 샘플링 기법은 데이터의 수를 조절하는 방법에 따라 소수클래스의 데이터를 다수클래스의 데이터만큼 증폭시키는 오버샘플링(Oversampling)과 다수클래스의 데이터를 소수클래스에 맞추어 감소시키는 언더샘플링(Undersampling)이 있다. 오버샘플링의 대표적인 기법으로는 재표집(Resampling), SMOTE (Synthetic Minority Oversampling Technique), Borderline-SMOTE 등이 있으며, 언더샘플링에는 랜덤 언더샘플링, Tomek links 등이 있다. 하지만 언급한 샘플링 기법 중 언더샘플링 기법은 데이터를 제거하는 방식이기 때문에 정보의 손실이 초래된다는 문제가 있고, 오버샘플링 기법은 합성된 데이터가 기존 데이터와 유사하기 때문에, 유사한 데이터를 반복 학습함으로써 과대적합(Overfitting)이 야기될 수 있다는 문제가 있다.

알고리즘 수준의 방법은 데이터 수준의 방법에 비해 많은 연구는 진행되지 않고 있지만, 대표적인 방법으로는 비용민감학습이 있다. 비용민감학습은 클래스를 오분류했을 때 비용을 고려해 주는 것으로, 소수클래스의 비용함수(Cost function)에 높은 가중치를 부여하는 방식이며, 샘플링 기법의 단점을 보완하면서도 이미지와 같은 비정형데이터에 적용하기 쉬운 장점이 있어 본 연구에서 적용시킨 방법이다.

## 2.2 선행된 유사연구

국방분야에서 딥러닝을 활용하여 이미지를 인식하기 위한 시도는 다양하게 적용되고 있다. 하지만, 관련된 연구들은 대부분 데이터 균형을 전제로 하고 있다.

Jeong et al.<sup>[4]</sup>은 잠수함의 비음향 탐지율 향상을 위해 전술적인 상황에서 함정의 TV카메라로 획득한 영상을 기반으로 잠망경을 비롯한 마스트를 올리고 항해하는 잠수함을 탐지하는 것을 주목표로 잠수함의 마스트와 항해중 유사하게 보일 수 있는 표적들을 분류하는 연구를 진행하였다.

Choi et al.<sup>[2]</sup>은 야간 및 저시정 상황 하에서 함정의

EOTS(Electro Optical Tracking System)나 IRST(Infra-Red Search and Track)를 활용하여 획득한 영상으로 군함과 비군함의 분류를 용이하게 하기 위해 위 장비로 획득한 영상자료와 유사하게 이미지를 전처리한 후 학습하여 분류율이 향상됨을 확인하였다.

Lee et al.<sup>[3]</sup>은 전시에 적합이 선박자동식별시스템(Automatic identification system)을 의도적으로 끄는 경우와 같은 상황에서 함정이 획득한 영상으로 주변 선박과 군함을 분류하고, 단계적으로 군함 내에서도 세부적으로 종류를 나누는 시스템을 제안하였다. 하지만 위 연구들은 모두 데이터의 균형을 전제하고 있기 때문에 현실적인 문제에서는 데이터불균형에 의한 예측 성능 저하가 야기될 수 있다.

데이터불균형을 해결하기 위해서는 많은 연구가 진행되었다. Jeong et al.<sup>[9]</sup>은 생명보험사 고객데이터의 불균형 문제를 해결하기 위하여 불균형 비율에 따른 오버샘플링의 효과에 대한 연구를 진행하였다. Lee et al.<sup>[10]</sup>은 데이터의 불균형 문제를 해결하기 위하여 언더샘플링 하는 과정에서 모집단 분포를 효율적으로 추출하여 샘플링하는 방법을 연구하였다. 하지만 이러한 연구들은 모두 정형데이터를 활용한 데이터불균형 해결 노력의 일환이며, 이미지와 같은 비정형데이터의 불균형에 대한 연구는 극소수이다. 이미지의 데이터불균형을 해결하기 위해 비용민감학습을 적용한 예로 H. C. Kim.<sup>[11]</sup>은 클래스의 데이터 개수에 따라 가중치를 생성하는 적대적 비용민감학습으로 개와 고양이의 이미지의 불균형데이터 분류율이 향상됨을 보였다.

한편, 국방분야에서는 이미지 뿐만 아니라 불균형데이터를 처리하기 위한 연구 자체가 부족한데, 그 중 Kim et al.<sup>[12]</sup>은 부사관 진급 요인을 분석함에 있어 발생하는 불균형한 데이터를 처리하기 위해 샘플링 기법을 활용하였다.

본 연구에서는 국방분야에는 최초로 이미지 인식의 불균형데이터 문제를 개선하기 위한 방안에 대해 연구하였으며, 비용민감학습을 적용한 CNN 모델을 이용하여 잠수함/비잠수함 데이터세트와 군함/비군함 데이터세트를 학습함으로써 모델 성능이 향상될 수 있음을 보이고자 한다.

## 3. 제안하는 비용민감학습 기반 이미지 분류모델

본 연구는 국방분야의 이미지 불균형데이터 문제를

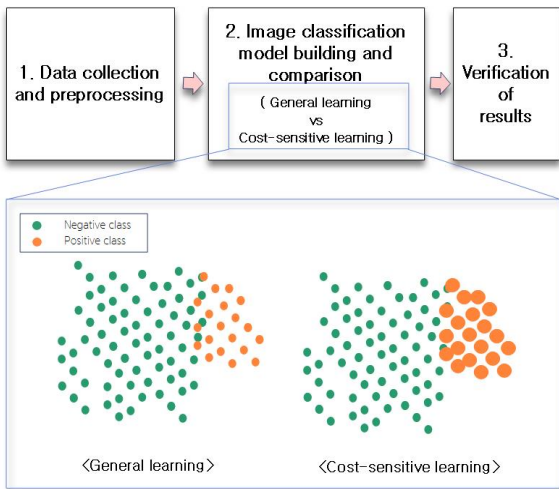


Fig. 2. A process of proposed classification model building

개선하기 위해 비용민감학습 기반 이미지 분류 모델 구축과정을 Fig. 2와 같이 데이터 수집 및 전처리, 이미지 분류모델 구축 및 비교 실험, 결과 검증의 3단계로 수행하였으며, 비용민감학습 기반 이미지 분류 모델과 일반 모델의 차이를 직관적으로 이해하기 쉽도록 그림으로 제시하였다.

1단계, 데이터 수집 및 전처리 단계에서는 잠수함/비잠수함과 군함/비군함 데이터를 수집하고, 불균형 데이터셋을 구성하였다. 2단계, 이미지 분류모델 구축 및 비교 실험은 2가지 데이터로 불균형 비율을 달리하여 구성된 총 4개의 데이터셋을 이용하여, 비용민감학습 기반의 다수 CNN 모델과 일반적인 다수 CNN 모델을 구축하였다. 이 후 각각의 모델로 테스트 이미지에 대한 예측 성능을 평가하였다. 3단계, 결과 검증 단계에서는 t-검정을 이용하여 비용민감학습 기반 CNN 모델 예측 성능의 평균값과 일반 CNN 모델 예측 성능의 평균값이 통계적으로 차이가 있는지를 검증하였다.

### 3.1 데이터 수집 및 전처리

본 연구에서 사용한 두가지 데이터에 대한 설명은 다음과 같다.

잠수함/비잠수함 데이터는 전술적인 상황에서 잠망경을 비롯한 마스트를 올리고 항해하는 잠수함을 탐지하는 것을 학습시키기 위해 사용한다. 이를 위해 마스트를 올리고 항해하는 잠수함, 부상한 잠수함 등의

잠수함 데이터와 수상함에서 잠수함 마스트로 오인할 수 있는 깃대부이, 등부표, 어망 부이, 소형 어선, 돌고래 지느러미 등 다양한 비잠수함 데이터를 구글 및 네이버 이미지를 검색하여 수집하였다. 이때, 현실에서 불균형이 발생할 것을 고려하여 잠수함 882장, 비잠수함 4,010장 총 4,892장을 수집하였다.

군함/비군함 데이터는 함정이 획득한 영상으로 주변 선박과 군함을 분류하는 것을 학습시키기 위해 사용하며, 선박자동식별시스템을 보완하기 위한 수단으로 활용하는 것이 목적이다. 이를 위해 군함, 여객선, 유조선, LNG선 등의 이미지를 캐글(Kaggle)을 통해 확보 하였고, 확보한 이미지로 구성할 수 있는 불균형 비율의 데이터 수량을 남긴 나머지는 임의로 삭제하였다. 따라서 군함 758장, 비군함 6,675장 총 7,433장을 수집하였다. 수집한 잠수함/비잠수함, 군함/비군함 데이터의 예는 Fig. 3과 같다.

수집한 잠수함/비잠수함 데이터는 1(Positive) : 5(Negative)의 불균형 비율을 가지고 있고, 불균형 비율에 따른 성능을 확인하기 위해 이 중 잠수함 데이터를 랜덤하게 삭제하여 1(Positive) : 10(Negative)의 불균형 비율을 가진 데이터셋을 추가로 구성하였다. 수집한 군함/비군함 데이터는 1(Positive) : 10(Negative)의 불균형 비율을 가지고 있었으나 마찬가지로 군함 데이터를 랜덤하게 삭제하여 1(Positive) : 20(Negative)의 불균형 비율 데이터셋을 추가로 구성하였다. 잠수함/비잠수함 데이터와 군함/비군함 데이터의 불균형 비율을 다르게 한 이유는 잠수함/비잠수함 데이터는 적절한 수량의 잠수함 데이터를 유지하면서 비잠수함 데이터 비율을 20까지 높일만큼 충분한 데이터를 확보하는 것이 제한되었기 때문이다.

훈련데이터(Training data)와 검증데이터(Validation data)는 일반적으로 7:3, 8:2, 9:1의 비율로 나눌 수 있는데, 본 연구에서는 많이 사용되는 8:2의 비율로 나누었고, 테스트데이터(Test data)는 테스트 정확도의 정확성을 위하여 별도로 100개의 데이터를 균일하게 가지도록 구성하였다. 세부 데이터 구성은 Table 1과 같다.

### 3.2 비용민감학습 기반 이미지 분류모델 구축 및 비교 실험

#### 3.2.1 비용민감학습 기반 이미지 분류모델 구축

본 연구에서는 이미지를 학습시키기 위해 사전 학습된(Pre-trained) CNN 모델을 전이학습시키는 과정을 거쳤다.



Fig. 3. Example images of data set(Google, Kaggle)

Table 1. Data set information

Class		Train- ing	Valida- tion	Test	Total	
Sub / Non-sub	1: 5	Sub	626	156	100	882
		Non-sub	3,128	782	100	4,010
	1: 10	Sub	313	78	100	491
		Non-sub	3,128	782	100	4,010
Warship / Non-warship	1: 10	Warship	526	132	100	758
		Non-Warship	5,260	1,315	100	6,675
	1: 20	Warship	263	66	100	429
		Non-Warship	5,260	1,315	100	6,675

CNN 모델은 이미지를 학습시키는데 최적화되어있는 딥러닝 알고리즘의 한 방법으로서 크게 컨볼루션층(Convolution layer)과 풀링층(Pooling layer), 완전연결층(Fully-connected layer)으로 구성된다<sup>[13]</sup>. 컨볼루션층은 입력이미지와 필터(Filter)와의 컨볼루션 연산을 통해 이미지의 특징을 추출하는 층으로, 학습을 통해 정해져야 하는 파라미터(Parameter)인 가중치(Weight)와 바이어스(Bias)가 정해진다. 풀링층(Pooling layer)은 이미지의 차원을 축소하여 필요한 연산량을 감소시키고, 이미지의 강한 특징만을 선별하는 층으로 파라미터를 갖지 않는다는 특성이 있다. 완전연결층(Fully-connected layer)은 컨볼루션층의 3차원 출력값을 1차원 벡터로 만들며, 모든 유닛들이 이전 층의 유닛들과 연결되어 있어 다수의 파라미터를 가지는 층이다. 가장 마지막의 완전연결층은 출력층의 역할을 하며, 출력층은 활성화함수를 거친 출력값에 따라 이미지의 클래스를 예측한다.

사전 학습된 CNN 모델이란 대량의 데이터셋을 이용해 미리 훈련되어 있는 CNN 네트워크로, 일반적으로 ILSVRC(Imagenet Large Scale Visual Recognition Competition)에서 100만장이 넘는 이미지 데이터셋인 이미지넷(Imagenet)을 이용하여 학습한 모델이 사용된다. 이 모델에서 학습된 가중치와 바이어스와 같은 파라미터와 네트워크의 구조를 이용하면 네트워크의 마지막 층 몇 개만 재학습하는 것만으로도 높은 성능을 낼 수 있는데, 이러한 학습 방법을 전이학습이라 한다.

본 연구에서 전이학습을 위해 사용한 사전 학습된 모델은 VGG16<sup>[14]</sup>이다. VGG16은 2014년 ILSVRC에서 준우승한 모델로 다른 모델과 대비하여 구조가 간단하고 성능이 우수하여 이미지의 전이학습 모델로 많이 사용된다. Fig. 4는 VGG16의 구조를 도식화해서 보여주는데, 13개의 컨볼루션층과 5개의 풀링층, 3개의 완전연결층으로 구성된 것을 알 수 있다.

본 연구에서 제시한 비용민감학습 기반 이미지 분류모델은 일반 CNN 모델과는 손실함수(Cost function)를 다르게 적용시킨다는 특징이 있다. 위의 CNN 모델을 포함한 딥러닝 모델에서는 손실함수를 정의하고 이를 최소화하는 방향으로 개선한다. 손실함수를 정의하기 위해서는 우선 데이터를 가중치와 바이어스를 포함한 가설함수로 나타낸 후 가설함수가 실제데이터와 얼마나 다른지를 나타내는데 이를 손실함수라 하며, 가중치와 바이어스를 업데이트 하면서 손실함수를 최소화시킨다.

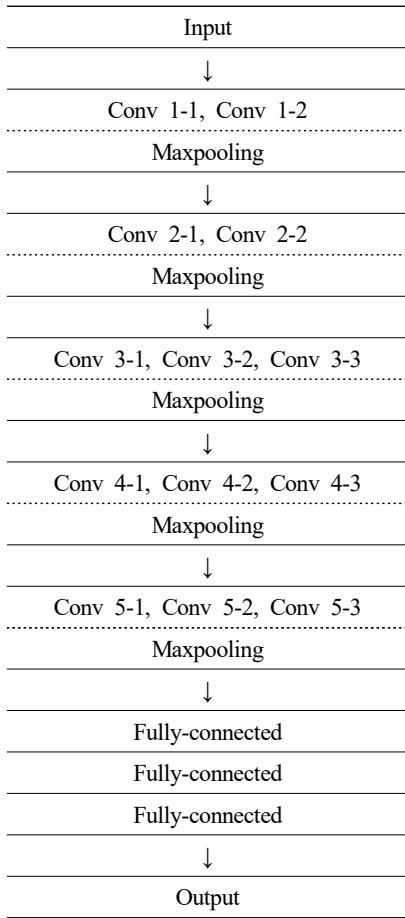


Fig. 4. A diagram of VGG16 structure

이미지 분류문제에서 주로 사용되는 손실함수 중 클래스가 2개 일 때 적용하는 이진 교차 엔트로피 함수(Binary cross entropy function)를 데이터 하나에 대해 나타내면 (1)과 같이 정의될 수 있다.

$$E = - \sum_{k=1}^K (t_k \log y_k + (1-t_k) \log (1-y_k)) \quad (1)$$

여기서  $t_k$ 는 정답 레이블,  $y_k$ 는 모델의 출력,  $k$ 는 데이터의 차원,  $K$ 는 데이터의 전체 차원 수를 의미한다. 이때 데이터가  $N$ 개라면 식은 (2)와 같이 나타낼 수 있다. 이 식이 의미하는 바는 각 데이터의 교차 엔트로피 함수를  $N$ 개의 데이터로 확장한 후, 다시  $N$ 으로 나눔으로써 정규화한 것으로<sup>[15]</sup>, 일반적인 이진 분류문제의 손실함수로 사용하는 식이다.

$$E = - \frac{1}{N} \sum_{n=1}^N \sum_{k=1}^K (t_{nk} \log y_{nk} + (1-t_{nk}) \log (1-y_{nk})) \quad (2)$$

(2) 식을 클래스 별로 교차 엔트로피 함수를 구하고 각 클래스의 교차 엔트로피 함수를 더해주는 방식으로 정의하면 (3)과 같이 나타낼 수 있다.

$$E = - \frac{1}{N} \sum_{j=1}^2 \sum_{m=1}^{M_j} \sum_{k=1}^K (t_{jmk} \log y_{jmk} + (1-t_{jmk}) \log (1-y_{jmk})) \quad (3)$$

여기서  $t_{jmk}$ 는 정답 레이블,  $y_{jmk}$ 는 모델의 출력,  $t_{jmk}$ 는 데이터의 차원,  $K$ 는 데이터의 전체 차원 수,  $m$ 은  $j$ 클래스의  $m$ 번째 데이터,  $M_j$ 은  $j$ 클래스가 갖는 데이터의 수,  $N$ 은 전체 데이터의 수를 의미한다.

본 연구에서는 사용한 비용민감학습의 손실함수는 클래스 별로 가중치를 다르게 적용한 방법으로 (4)와 같이 정의될 수 있다.

$$E = - \frac{1}{N} \left( \sum_{m=1}^{M_0} \sum_{k=1}^K (t_{jmk} \log y_{jmk} + (1-t_{jmk}) \log (1-y_{jmk})) \right) + C \sum_{m=1}^{M_1} \sum_{k=1}^K (t_{jmk} \log y_{jmk} + (1-t_{jmk}) \log (1-y_{jmk})) \quad (4)$$

여기서  $M_0$ 은 다수클래스의 데이터 수,  $M_1$ 은 소수클래스의 데이터 수,  $C$ 는 클래스 가중치를 의미한다. 즉 소수클래스를 오분류 했을 경우에는  $C$ 배만큼 큰 손실함수 값을 갖게 됨을 의미한다. 여기서 정답 레이블인  $t_{jmk}$ 은 0또는 1 중 하나의 값을 갖게 되며, 다수클래스의  $t_{jmk}$ 은 0, 소수클래스의  $t_{jmk}$ 은 1이 된다. 따라서 (4)는 (5)와 같이 다시 정의할 수 있다.

$$E = - \frac{1}{N} \left( \sum_{m=1}^{M_0} \sum_{k=1}^K (1-t_{jmk}) \log (1-y_{jmk}) \right) + C \sum_{m=1}^{M_1} \sum_{k=1}^K t_{jmk} \log y_{jmk} \quad (5)$$

클래스 가중치  $C$ 를 결정하는 방법에 대해 일반화된 이론은 아직 알려지지 않았으며, 통상 불균형 비율을 역으로 반영하여 균형을 이루게 하는 방법이 많이 활용되고 있다<sup>[16]</sup>.

3.2.2 비교 실험

비교 실험을 위해 우선적으로 하이퍼파라미터를 설정하였다. 하이퍼파라미터란 연구자가 지정해 주어야 하는 다양한 값으로 네트워크 구조의 수정 정도, 얼마나 많은 층을 재학습 시킬 것인지 여부와 완전연결층의 유닛(Unit)의 개수, 활성화 함수, 학습세대(Epoch), 배치(Batch), 학습률(Learning rate) 등이 그 예이다. 본 연구에서는 VGG16의 네트워크 구조 중 마지막 완전연결 3개층을 삭제하고 새로운 1개층을 구성한 뒤 완전연결층의 유닛 개수, 학습률, 배치 등을 변경해가며 학습결과를 확인하였다. 하이퍼파라미터 튜닝의 세부 값은 Table 2와 같다. 또한, 동일한 하이퍼파라미터 설정 조건을 가진다고 하더라도 가중치 초기값이 매번 달라지면서 실험 결과의 오차가 조금씩 발생할 수 있기 때문에 동일 조건에서 3~4번씩 학습시켰다. 즉 하나의 데이터세트에서 하이퍼파라미터의 설정이 다른 6가지 경우로 각각 3~4번 학습을 하되, 이를 비용민감 학습 기반 모델에서 20회 학습하고 일반 모델에서 20회 학습하였다. 동일한 실험을 총 4가지 데이터세트에서 실시하였기 때문에 160회의 실험 후 각각의 결과를 확인하였다. 세부적인 실험 과정은 Fig. 5와 같다. 이 중 비용민감학습 시 소수클래스에 대한 가중치는 사전분석을 통해 불균형 비율의 역이나 그 이상일 때 성능이 향상됨을 확인하고, 이를 반영하여 불균형 비율을 역으로 반영하는 방법을 사용하였다. 즉 불균형 비율이 1(Positive) : 10(Negative)인 경우 소수클래스인 양성에 10의 가중치를 적용하였다.

모델 성능을 측정하기 위해 확인한 성능평가 지표는 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1-score이며, 이 중 F1-score을 기준으로 모델 성능을 평가하였다. 각 성능평가 지표들이 의미하는 바는 Table 3과 같은 정오분류표(Confusion matrix)를 이용해 나타낼 수 있는데, 그 의미는 아래와 같다.

Table 2. Hyperparameter tuning

Fully-connected layer number		Fully-connected layer unit number		Activation function
1		256, 512, 1024		relu
Drop out ratio(%)	Epoch	Batch size	Learning rate	
0.5	200	32, 100	10 <sup>-4</sup> , 10 <sup>-5</sup>	

Table 3. Confusion matrix

-		True condition	
		Positive	Negative
Predicted condition	Positive	True positive (TP)	False positive (FP)
	Negative	False negative (FN)	True negative (TN)

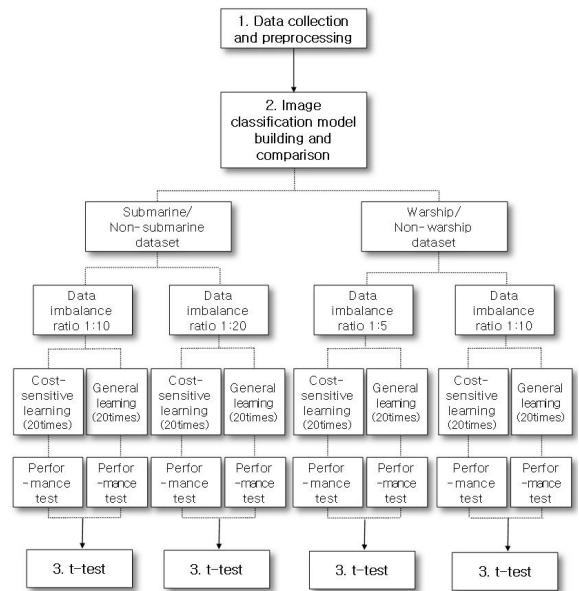


Fig. 5. A detail process of experiments

일반적으로 머신러닝이나 딥러닝의 분류모델에서 성능평가를 위해 가장 많이 사용하는 지표는 정확도(Accuracy)이다. 정확도는 (6)과 같이 정의할 수 있다.

$$ACC = \frac{TP + FN}{TP + FP + TN + FN} \tag{6}$$

그러나 정확도는 불균형데이터에서 단일 성능평가 지표로 사용되기에는 제한사항이 있다. 90개의 음성 데이터와 10개의 양성인 데이터가 있는 상황을 생각해 볼때, 만약 100개의 데이터 전체를 음성으로 분류를 하는 모델이 있다면 정확도는 90 %라고 할 수 있다. 하지만, 세부적으로 살펴보면 이 모델이 양성인 데이터를 분류해 낼 가능성은 0 %인 것이다. 따라서 불균형데이터에서는 정밀도(Precision), 재현율(Recall), F1-

score와 같은 성능평가 지표를 추가적으로 확인하여야 한다.

정밀도(Precision)는 모델이 양성(Positive class)으로 검출해낸 것 중 실제 양성인 것을 의미하는 기준으로 (7)과 같이 나타낼 수 있다.

$$Precision = \frac{TP}{TP+FP} \tag{7}$$

재현율(Recall)은 실제 해당 데이터 중 모델이 올바르게 양성으로 검출해낸 것을 의미하는 기준으로 양성인 것을 놓치지 않고 식별해 내는지를 판단하는 지표이며 (8)과 같다.

$$Recall = \frac{TP}{TP+FN} \tag{8}$$

정밀도와 재현율은 한 수치가 높아지면 다른 수치가 낮아지는 관계를 가지고 있기 때문에, 정밀도와 재현율 중 하나의 수치가 극단적으로 높은 것보다는 두 수치가 적절하게 조화를 이루는 모델이 더 좋은 모델이라고 할 수 있다<sup>[7]</sup>. 이를 위한 평가 지표인 F1-score는 (9)와 같이 정밀도와 재현율의 조화평균으로 나타낼 수 있다.

$$F1\text{-score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{9}$$

Table 4. F1-score results for each model

Sub / Non-sub					Warship / Non-warship				
Hyper parameter	1 (positive) : 5 (negative)		1 (positive) : 10 (negative)		Hyper parameter	1 (positive) : 10 (negative)		1 (positive) : 20 (negative)	
	Cost-sensitive learning	General learning	Cost-sensitive learning	General learning		Cost-sensitive learning	General learning	Cost-sensitive learning	General learning
Learning rate $10^{-4}$ , Unit 512, Batch 100	72.15	58.74	64.90	50.00	Learning rate $10^{-4}$ , Unit 512, Batch 32	85.23	85.88	81.18	78.79
	72.05	58.74	67.94	52.17		86.52	83.04	75.00	66.67
	72.50	52.17	70.44	41.27		88.40	87.78	81.87	75.00
Learning rate $10^{-5}$ , Unit 512, Batch 32	73.42	70.13	70.89	52.17	Learning rate $10^{-5}$ , Unit 512, Batch 32	88.40	87.78	77.58	85.88
	65.77	47.33	70.89	54.29		87.15	84.57	84.57	78.00
	68.42	52.17	64.47	51.47		87.78	80.47	77.58	77.54
Learning rate $10^{-4}$ , Unit 256, Batch 32	73.76	73.29	72.05	45.45	Learning rate $10^{-4}$ , Unit 256, Batch 32	91.40	91.98	78.31	76.83
	74.07	58.16	72.39	46.62		90.22	89.01	79.62	75.31
	78.57	70.44	74.84	51.09		90.22	86.52	82.56	66.23
Learning rate $10^{-4}$ , Unit 256, Batch 100	76.36	73.62	70.44	43.07	Learning rate $10^{-4}$ , Unit 256, Batch 100	93.12	89.62	83.24	74.53
	63.09	64.00	72.50	44.27		90.32	89.62	87.15	73.75
	73.75	69.23	69.62	51.47		89.62	89.62	87.29	79.04
Learning rate $10^{-4}$ , Unit 1024, Batch 100	70.44	55.00	67.97	55.71	Learning rate $10^{-4}$ , Unit 1024, Batch 100	89.62	92.55	84.57	81.18
	73.29	55.94	71.20	36.00		89.62	90.21	81.18	79.76
	60.69	50.75	73.62	44.27		90.81	89.01	87.78	85.19
	68.83	58.74	70.06	39.37		90.81	90.81	69.68	83.24
Learning rate $10^{-5}$ , Unit 1024, Batch 100	67.11	51.85	72.50	44.62	Learning rate $10^{-5}$ , Unit 1024, Batch 100	92.63	85.23	85.23	73.75
	76.07	48.89	69.62	51.47		91.4	87.15	81.87	74.85
	67.97	54.29	69.23	54.29		93.12	87.15	82.56	79.04
	61.64	45.45	72.05	52.17		90.32	90.22	81.87	71.34
Mean	<b>70.50</b>	<b>58.45</b>	<b>70.38</b>	<b>48.06</b>	Mean	<b>89.84</b>	<b>87.91</b>	<b>81.53</b>	<b>76.80</b>
Std. Deviation	<b>4.96</b>	<b>8.82</b>	<b>2.61</b>	<b>5.51</b>	Std. Deviation	<b>2.11</b>	<b>3.02</b>	<b>4.41</b>	<b>5.21</b>



160회 실험의 F1-score 결과값과 통계량은 Table 4와 같이 잠수함/비잠수함 데이터셋에서 불균형 비율이 1(Positive) : 5(Negative)인 경우 비용민감학습 기반 이미지 분류모델의 F1-score 평균값이 70.50 %로 일반 모델의 58.45 % 보다 12.05 % 높았고, 1(Positive) : 10(Negative)인 경우에는 비용민감학습 기반 이미지 분류모델의 값이 70.38 %로 일반 모델보다 22.32 % 높았다. 군함/비군함 데이터셋에서 불균형 비율이 1(Positive) : 10(Negative)인 경우 비용민감학습 기반 이미지 분류모델의 F1-score 평균값이 89.84 %로 일반 모델의 87.91 % 보다 1.93 % 높았고, 1(Positive) : 20(Negative)인 경우는 비용민감학습 기반 이미지 분류모델의 값이 81.53 %로 일반 모델보다 4.73 % 높았다.

비용민감학습을 적용함에 따른 성능 개선은 불균형 비율이 높을수록 커짐을 알 수 있었다. 불균형 비율이 상대적으로 낮을 때 비용민감학습을 적용하지 않았을 때의 F1-score이 오히려 높은 경우가 일부 있었지만, 불균형 비율이 높을 때는 그 경우가 줄어들었다. 평균 또한 불균형 비율이 높을 때 더 크게 향상되었으며, 표준편차는 4개의 그룹에서 모두 줄어든 것을 확인할 수 있었다. 잠수함/비잠수함 데이터셋과 군함/비군함 데이터셋가 동일한 불균형 비율인 1(Positive) : 10(Negative)에서 비용민감학습 적용 여부에 따른 성능 개선 효과가 크게 차이 나는 것은 두 데이터셋의 데이터 수량 차이와 군함/비군함 데이터 세트의 표적 크기가 크고 특징이 상대적으로 뚜렷해 구별이 용이하기 때문인 것으로 판단하였다.

#### 4. 통계적 검증

최종적으로 비용민감학습 기반 이미지 분류모델의 성능이 통계적으로 유의미하게 우수하다는 것을 확인하기 위해 R프로그램을 이용하여 독립표본 t-검정을 실시하였다.

t-검정을 시행하기 위해서는 t-검정의 가정인 정규성(Normality)과 등분산성(Homoskedasticity)을 확인하여야 한다. 데이터가 정규성을 만족할 때에 t-검정을 실시할 수 있고, 등분산성을 확인하여 등분산일 경우 등분산 t-검정, 이분산(Heteroskedasticity)일 경우 이분산 t-검정을 시행한다.

#### 4.1 정규성 확인

정규성 확인은 비용민감학습 기반 이미지 분류모델로 평가한 20개의 F1-score값과 일반 모델로 평가한 20개의 F1-score값들의 각각의 정규성 여부를 평가한다. 본 연구에서는 4개의 데이터셋을 활용하였기 때문에 총 8개의 그룹에 대해 정규성을 확인하였다.

정규성은 시각적 방법과 검정을 통해 확인할 수 있다. 시각적으로는 Q-Q도를 통해 확인할 수 있는데, Q-Q도는 데이터가 특정 분포를 따르는지를 시각적으로 검토하는 방법이다. 여기서 Q는 분위수(Quantile)의 약어로, 분위수들을 차트에 그리고 나면 데이터와 비교하고자 하는 분위수 간에 직선 관계가 보이는지 확인할 수 있다<sup>[18]</sup>. Q-Q도를 활용하여 정규분포를 확인할 때에는 데이터가 45도 선에 가까울수록 정규성을 가진다고 말할 수 있는데, 본 연구에서 사용한 그룹별 20개 F1-score값을 Q-Q도로 나타낸 결과값들은 45도에 가깝게 분포함을 확인할 수 있었다.

그러나 시각적 확인만으로는 정규성을 가진다고 말할 수 있는지 여부가 불확실하다. 따라서 정규성 검정 방법 중 하나인 샤피로 윌크 검정(Shapiro-wilk's test)으로 확인해 보았다. 샤피로 윌크 검정의 귀무가설은 데이터가 정규분포라는 것이며, 대립가설은 정규분포와 다르다는 것이다.

샤피로 윌크 검정 결과 p-value값은 Table 5와 같이 통상적으로 유의수준으로 사용하는 0.1, 0.05, 0.01 보다 모두 큰 값으로 귀무가설을 기각할 수 없다. 따라서 각 데이터들은 정규성을 가진다고 말할 수 있다.

#### 4.2 등분산성 확인

등분산성은 비용민감학습 기반 이미지 분류모델로 평가한 F1-score값들과 일반 모델로 평가한 F1-score값들의 분산이 동일한지 여부를 평가한다. 등분산성을 검정 방법으로 확인하기에 앞서 상자그림(Boxplot)을 통해 데이터의 분포를 시각적으로 확인해 보았다. 상자그림은 데이터의 1:3사분위수, 중앙값(Median), 최대값, 최소값, 이상치 등을 나타내는 그림으로 상자 안에 1:3사분위수 내에 있는 데이터들이 존재한다. 시각상 잠수함/비잠수함 데이터로 학습시킨 모델의 테스트 F1-score값은 비용민감학습 기반 이미지 분류모델과 일반 모델의 평균 및 분산 차이가 비교적 크고, 군함/비군함 데이터로 학습시킨 모델은 평균은 비용민감학습 기반 이미지 분류모델의 결과가 높지만 분산은 크게 다르지 않음을 알 수 있었다.

하지만 시각만으로는 등분산성을 가지는지 여부를 정확하게 단정하기 어렵다. 따라서 추가적으로 두 그룹의 분산을 비교하는 F-검정을 통해 등분산성을 확인해 보았다. F-검정의 귀무가설은 두 집단의 데이터는 분산이 같다는 것이며, 대립가설은 두 집단의 데이터는 분산이 다르다는 것이다. 유의수준은 0.05로 지정하였다.

F-검정 결과 Table 6과 같이 잠수함/비잠수함 데이터로 학습시킨 모델의 p-value값은 0.05보다 작아 귀무가설을 기각, 대립가설을 채택할 수 있고, 군함/비군함 데이터로 학습시킨 모델들의 p-value값은 0.05보다 커 귀무가설을 기각할 수 없다. 따라서 잠수함/비잠수함 데이터로 학습시킨 모델은 비용민감학습 기반 이미지 분류모델의 테스트 F1-score값과 일반 모델의 테스트 F1-score값의 분산이 다르며, 군함/비군함 데이터로 학습시킨 모델은 두 그룹간 분산이 같다고 할 수 있다.

Table 5. Results of shapiro-wilk's test

Data set	Imbalance ratio	Cost-sensitive learning application	p-value
Sub / Non-sub	1 (positive) : 5 (negative)	Cost-sensitive learning	0.3593
		General learning	0.1056
	1 (positive) : 10 (negative)	Cost-sensitive learning	0.2427
		General learning	0.1252
Warship / Non-warship	1 (positive) : 10 (negative)	Cost-sensitive learning	0.6163
		General learning	0.4962
	1 (positive) : 20 (negative)	Cost-sensitive learning	0.2088
		General learning	0.6177

Table 6. Results of F-test

Data set	Imbalance ratio	p-value
Sub / Non-sub	1 (positive) : 5 (negative)	0.01564
	1 (positive) : 10 (negative)	0.002074
Warship / Non-warship	1 (positive) : 10 (negative)	0.1294
	1 (positive) : 20 (negative)	0.4766

### 4.3 t-검정

정규성 검정과 등분산성 검정 결과를 바탕으로, 2가지 데이터로 불균형 비율을 서로 달리하여 구성된 총 4개의 데이터셋을 이용하여 학습한 모델들의 F1-score값에 대한 t-검정을 실시하였다. 검정 내용은 각 데이터셋 별로 비용민감학습 기반 이미지 분류모델의 테스트 F1-score 평균값과 일반 모델의 테스트 F1-score 평균값이 통계적으로 유의미하게 다른지 확인하는 것이다. 이를 위한 가설 설정은 다음과 같다.

$$\text{귀무가설}(H_0) : \mu_{\text{cost-sensitive}} = \mu_{\text{general}}$$

$$\text{대립가설}(H_1) : \mu_{\text{cost-sensitive}} > \mu_{\text{general}}$$

이때  $\mu_{\text{cost-sensitive}}$  는 비용민감학습 기반 이미지 분류모델의 테스트 F1-score 평균값이고,  $\mu_{\text{general}}$  는 일반 모델의 테스트 F1-score 평균값이다. 유의수준은 0.05로 선정하였다.

검정 결과는 Table 7과 같이 4가지 데이터셋에서 모두 p-value값이 0.05보다 작아 귀무가설을 기각하고, 비용민감학습 기반 이미지 분류모델의 테스트 F1-score 평균값이 일반 모델의 테스트 F1-score 평균값보다 크다는 대립가설을 채택하였다. 따라서 비용민감학습 기반 이미지 분류모델의 성능이 우수함을 통계적으로 검증하였다.

Table 7. Results of t-test

Data set	Imbalance ratio	Test type	p-value
Sub / Non-sub	1 (positive) : 5 (negative)	Non-equal variance t-test	$4.69 \times 10^{-6}$
	1 (positive) : 10 (negative)		$6.8 \times 10^{-16}$
Warship / Non-warship	1 (positive) : 10 (negative)	Equal variance t-test	0.01275
	1 (positive) : 20 (negative)		0.001823

## 5. 결론

본 연구는 국방분야 이미지 분류의 불균형데이터 문제를 개선하기 위한 방안으로 잠수함/비잠수함 데이

터세트와 균함/비균함 데이터세트에 대해 클래스에 따라 서로 다른 비용을 부여하는 비용민감학습(Cost-sensitive learning)을 적용하여 CNN 모델을 통해 학습시켰다.

2가지 데이터로 불균형 비율을 서로 달리한 총 4개의 데이터세트를 이용하여 비용민감학습 기반 모델 80개, 일반 모델 80개를 구축 후 동일한 테스트데이터에 대한 예측 성능을 평가하였다.

결과적으로 4개의 데이터세트에서 비용민감학습 기반 이미지 분류모델의 F1-score 평균값이 일반 모델의 F1-score 평균값보다 최대 22.32 %에서 최소 1.93 % 높았다. 또한 이 결과가 통계적으로 유의미한지 확인하기 위해 정규성 검정과 등분산성 검정 결과를 바탕으로 t-검정을 실시결과 4가지 데이터세트에서 모두 p-value값이 0.05보다 작아 비용민감학습 기반 이미지 분류모델의 F1-score 평균값이 크다는 대립가설이 유의미함을 확인하였다.

본 연구의 제한사항은 다양한 불균형데이터 처리기법을 적용해 보지 못했다는 점이다. 클래스별 데이터 확장(Data augmentation), GAN(Generative Adversarial Network)을 통한 오버샘플링이나 이미지 삭제를 통한 언더샘플링 등의 방법과 비용민감학습의 효과도 비교를 진행하지 못하였다. 또한, 비용민감학습 중 본 연구에서 진행한 손실함수를 수정하는 방법 외 알고리즘에 관한 연구로 진행되고 있는 엔트로피(Entropy) 기반 SVM, 비용민감 랜덤 포레스트(Cost-sensitive random forest) 등과의 비교는 향후 추가 연구를 통해 보완해야 할 과제이다.

## References

- [1] S. W. Park, H. S. Park, Y. B. Kim, D. G. Kim, "Suspected Object Detection and Tracking methods for National Defence Surveillance," Communications of the Korean Institute of Information Scientists and Engineers, Vol. 36, No. 8, pp. 33-40, 2018.
- [2] D. W. Choi, J. M. Ma, "Improvement of Ship Classification Based on Artificial Neural Network Considering Night and Low Visibility Marine Environment," Korean Journal of Computational Design and Engineer-ing, Vol. 24, No. 3, pp. 320-328, 2019.
- [3] J. W. Lee, D. W. Nam, W. Y. Yoo, "A Study on Composition of Warship Type/Class Identification System," The Korea Institute of Information and Communication Engineering, Vol. 23, No. 2, pp. 607-610, 2019.
- [4] M. A. Jeong, J. M. Ma, "Improvement of Submarine Detection Based on Mast Images Using An Ensemble Model of Convolutional Neural Networks," Journal of the Korea Institute of Military Science and Technology, Vol. 23, No. 2, pp. 115-124, 2020.
- [5] Log0, "Class Imbalance Problem," August 30, 2013, <http://www.chioka.in/class-imbalance-problem/> (accessed July 10, 2020.).
- [6] H. He, E. A. Garcia, "Learning from Imbalanced Data," IEEE Transactions On Knowledge And Data Engineering, Vol. 21, No. 9, pp. 1263-1284, 2009.
- [7] S. B. Kim, "Sampling Technique for Analysis of Imbalanced Data," February 20, 2020, <https://www.youtube.com/watch?v=Vhwz228VrIk> (accessed June 28, 2020.).
- [8] H. S. Jeong, C. W. Kang, K. K. Kim, "The Effect of Oversampling Method for Imbalanced Data," Journal of The Korean Data Analysis Society, Vol. 10, No. 4, pp. 2089-2098, 2008.
- [9] K. N. Lee, J. T. Lim, K. S. Bok, J. S. Yoo, "Handling Method of Imbalance Data for Machine Learning : Focused on Sampling," The Korea Contents Society, Vol. 19, No. 11. pp. 567-577, 2019.
- [10] H. C. Kim, "Adversarial Cost-Sensitive Learning for Class Imbalance Problem," M. E. Thesis, University of Gangwon, 2020.
- [11] T. Fawcett, "Learning from Imbalanced Classes," August 25, 2016, <https://www.svds.com/learning-imbanced-classes/> (accessed July 10, 2020.).
- [12] J. O. Kim, E. S. Hwang, J. M. Yoo, "Analysis of NCO Promotion Factors Using Imbalanced Data Processing," 12<sup>th</sup> International Army Modeling & Simulation Education Conference, p. 90, 2019.
- [13] Solaris, "Deep Learning to Learn by Tensorflow," Youngjin.com, Seoul, 2018.
- [14] K. Simonyan, A. Zisserman, "Very Deep

- Convolutional Networks for Large-Scale Image Recognition,” International Conference on Learning Representations, pp. 1-14, 2015.
- [15] G. Saito, “Deep Learning from Scratch,” Hanbit Media, Seoul, 2017.
- [16] J. Brownlee, “How to Develop a Cost-Sensitive Neural Network for Imbalanced Classification,” February 2, 2020, <https://machinelearningmastery.com/cost-sensitive-neural-network-for-imbalanced-classification/> (accessed May 2, 2021.).
- [17] S. J. Yoo et al., “Analysis of Big-Data Using Python,” The Twenty-First Century, Paju, 2018.
- [18] M. K. Seo, “Practical Data Processing and Analysis Using R,” Gilbut Publishing, Seoul, 2014.