

## 캔들스틱 차트와 합성곱 신경망을 이용한 주가 갭 예측 연구

김찬수, 최흥식, 김선웅\*  
국민대학교 비즈니스IT전문대학원

### Price gap prediction using candlestick chart and convolutional neural network

Chansu Kim, Heung Sik Choi, Sun Woong Kim\*  
Graduate School of Business IT, Kookmin University

**요약** 기술적 분석에서 가장 활용도가 높은 캔들스틱 차트는 주가의 추세를 직관적으로 파악할 수 있는 도표이다. 캔들스틱 차트 분석으로 수익을 얻을 수 있다면 딥러닝 알고리즘은 학습을 통해 주가 움직임 패턴을 찾아낼 수 있을 것이다. 본 연구의 목적은 주가의 일중, 일별 캔들스틱 차트의 이미지를 CNN에 학습시킨 후, 전일 종가와 당일 시가 차이로 측정하는 주가 갭을 예측하는 것이다. KOSPI200을 구성하는 개별종목을 대상으로 실증 분석한 결과는 다음과 같다. 첫째, 제안 CNN 모형은 갭 상승과 하락을 예측하는 이진 분류에서 55%의 정확도를 보였다. 둘째, 갭 상승과 하락을 나누어 분석한 결과 갭 상승에서 58%, 갭 하락에서 54%의 정밀도를 보였다. 셋째, 표본 외 자료로 시뮬레이션한 결과 제안 모형은 벤치마크 모형보다 우수한 예측 성과를 보였다. 본 연구는 이미지에 다양한 캔들스틱 차트 정보를 담은 CNN 모형에 학습시킨 점과 단기거래에서 중요한 역할을 하는 주가 갭을 예측했다는 점에서 기존 연구와 차별점이 있다. 향후 연구에서는 투자 주체별 거래 정보를 이미지에 추가하여 예측 성과를 개선할 필요가 있다.

**Abstract** The candlestick chart is a diagram that can be used for the intuitive capture of stock price trends. Hence, if profits can be made by the implementation of candlestick chart analysis, deep learning algorithms would be able to determine stock price movement patterns. The purpose of this study was to predict the price gap measured by the difference between the previous day's closing price and today's opening price after the assessment of images of intraday and daily candlestick charts using the CNN model. The following empirical results were obtained for individual stocks in the KOSPI 200 index. First, the proposed CNN model had 55% accuracy in the binary classification predicting the up and down gap. Second, an accuracy of 58% was achieved in the upward move of the gap and 54% in its downward move. Third, the simulation based on out-of-sample data increased the CNN gap prediction model performance, which was higher than that of the benchmark model. In future studies, the predictive performance should be improved by the addition of transaction information of major investor groups to the images.

**Keywords** : Trading System, CNN, Deep Learning, Price Gap, Candlestick Chart

---

\*Corresponding Author : Sun Woong Kim(Kookmin Univ.)

email: swkim@kookmin.ac.kr

Received November 18, 2022

Accepted January 6, 2023

Revised January 4, 2023

Published January 31, 2023

## 1. 서론

알고리즘 매매(algorithm trading)는 감정을 배제하고 원칙대로만 거래하기 때문에 인간이 가지고 있는 행태적 편의(behavioral bias)를 극복할 수 있다는 장점이 있다. 반면 진화하는 시장에서 살아남기 위해서는 끊임 없이 아이디어를 발굴하고 활용해야 하는 수고가 따른다. 이는 투자자의 직관이 필요한 부분이다.

인공지능(artificial intelligence)은 이러한 문제를 해결할 수 있는 대안으로 떠올랐다. 데이터만 주면 스스로 거래전략을 학습하고, 시장의 변화에 따라 진화하며, 감정의 휘둘림 없이 거래하는 이상적인 방식이 가능해진 것이다. 이에 인공지능을 이용한 투자전략에서는 크게 두 가지 영역에서 꾸준한 연구가 이루어지고 있다.

첫 번째는 금융 데이터와 인공지능 모형을 이용하여 주가를 예측하는 연구이다. 예를 들어 거래의 흐름을 파악하여 최적화된 매수, 매도 시점을 파악하거나[1], 주가 움직임의 패턴으로 방향을 예측하는 것이다[2,3].

두 번째는 기존에 활용하지 않았던 대체 자료(alternative data)를 이용한 연구이다. 뉴스의 감성 분석을 통해 시장의 심리(sentiment)를 파악하거나[4], 포털 사이트의 검색지수를 이용한 주가 예측 등이 여기에 해당한다[5].

본 연구의 목적은 실제 투자에서 활용되고 있는 캔들스틱 차트(candlestick chart) 패턴과 딥러닝 모형을 결합하여 주가의 갭(gap)을 예측하는 것이다. 특히 대부분의 딥러닝 모형이 1차원 데이터를 이용하여 시계열 자료를 예측했던 것과 달리, 합성곱 신경망(CNN : Convolutional Neural Network, 이하 CNN)을 이용하여 2차원의 이미지 데이터를 학습하고 주가 갭을 예측한다. 2차원 이미지 데이터로 주가 예측을 시도하는 이유는 투자자가 숫자가 아닌 캔들스틱 차트의 패턴을 보면서 거래하기 때문이다. 투자자가 이미지의 패턴을 감지할 수 있다면 딥러닝 알고리즘도 패턴을 찾아낼 수 있을 것이다.

본 연구의 방법은 다음과 같다. 첫째, 2017년 1월부터 2022년 8월까지의 기간에서 KOSPI200을 구성하는 개별종목의 일중 주가 자료와 KOSPI, S&P500, NASDAQ 지수의 일별 자료로 이미지를 생성하고, 이미지에 갭 수익률을 라벨링(labeling) 한다.

둘째, 갭 상승과 갭 하락을 예측하는 이진 분류 CNN 모형을 설계하고 이미지를 모형에 학습시킨 후, 모형의 성능을 평가할 것이다.

셋째, 학습, 검증, 테스트에 이용되지 않은 표본 외 자료를 예측하여 결과를 평가하고, 예측 결과값으로 룬숫

전략(long-short strategy, L-S)을 시뮬레이션하여 실제 투자에서의 활용 가능성을 살펴볼 것이다.

## 2. 이론적 배경

### 2.1 캔들스틱 차트와 주가 패턴

단기거래자 대부분이 실패를 경험하지만, 일부는 캔들스틱 차트를 보고 직관적으로 주가 패턴을 파악한 후 거래를 통해 수익을 내고 있다. Fig. 1은 단기거래자가 기술적 분석(technical analysis)에서 일반적으로 사용하는 캔들스틱 차트를 나타낸 그림이다. 캔들의 몸체(body)는 시가와 종가를 나타내며, 시가보다 종가가 높으면 빨간색, 낮으면 파란색으로 표시한다. 위아래의 꼬리(tail)는 고가와 저가를 표시한다. Fig. 2는 캔들스틱 차트의 시계열 자료에 대한 패턴의 예시를 보여주고 있다.

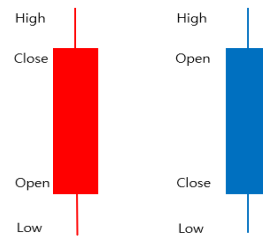


Fig. 1. Composition of candlestick

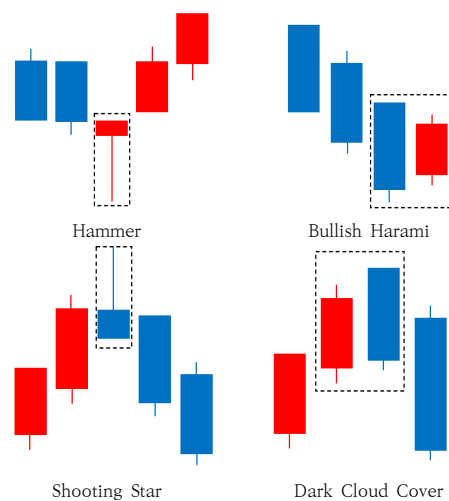


Fig. 2. Example of candlestick pattern

주가 패턴 연구는 투자 성과에 직결되는 만큼 그동안 많은 연구가 이루어졌다. Lee et al.(1997)은 주시가격의 변동을 예고하는 패턴들을 정의하고 이를 이용해 차트해석기(chart interpreter)를 개발하였다[6]. 1992년부터 1997년까지 데이터를 이용하여 패턴 매매 방식을 시뮬레이션한 결과 평균 투자 성공률이 72%를 보였지만, 사용자가 패턴을 정의해주어야 한다는 부분에 한계가 있다고 밝혔다. 이후 패턴을 이용한 분석은 Lee(2012)의 이동평균선 패턴과 전환점 행렬을 이용한 연구, Yang et al.(2019)의 M&W 패턴과 유전자 알고리즘을 이용한 연구처럼 지능형 매매시스템을 이용한 연구로 발전하였다[7,8].

### 2.2 주가 갭

거래가 중단되는 전일의 장 마감 시점부터 당일 장 시작 시점까지 거래가 중단되는 시간 동안 새로운 정보가 발생한다면 당일 시가는 전일 종가와 차이가 발생할 수 있다. Fig. 3은 전일 야간 시간에 발생한 호재에 의해 당일 시가가 전일 종가보다 상승으로 시작하는 상승 갭을 보여주고 있다.



Fig. 3. Example of price gap

주가 갭은 식 (1)과 같이 정의한다.

$$gap_t = \frac{Open_t}{Close_{t-1}} - 1, \tag{1}$$

where  $Open_t$  is opening price on day  $t$ ,  
 $Close_{t-1}$  is closing price on day  $t-1$ .

주가 갭 이례 현상(price gap anomaly)에 관한 초기 연구는 요일 효과(the day of the week effect)를 중심으로 이루어졌다. Cross(1973)는 금요일과 월요일 사이에 주가 수익률 분포가 변함을 밝혔으며[9], French(1980), Gibbons and Hess(1981), Keim and Stambaugh (1984) 등에 의해 주말효과(weekend effect)가 연구되

었다[10-12].

주가 갭에 관한 다른 연구로는 시가가 형성될 때 주가가 과잉반응하고 장중에는 과잉반응이 해소된다 주장하는 과잉반응가설(overreaction hypothesis)이 있다. Grant et al.(2005)은 시가에 뚜렷한 양의 수익률이 나타나고 이후 장중에는 수익률이 반전한다고 분석했다[13].

최근 문헌 중 Ham et al.(2022)은 미국 주식시장에서 야간수익률이 주간수익률보다 높은 오버나잇 퍼즐(overnight puzzle) 현상이 나타남을 밝혔다[14]. Kim(2022)은 이 현상을 코스피 주가지수에 적용하여 실증 분석한 결과, 2003년 2월부터 2022년 6월까지의 기간에서 야간수익률의 평균이 하루 전체 수익률의 평균보다 4.08배 이상 높게 나타난다 분석했다[15].

### 2.3 합성곱 신경망과 주가 예측

CNN은 LeCun et al.(1989)이 문자인식을 위해 소개한 방법으로, 픽셀로 변환한 이미지의 특징을 추출하여 분류(classification)하는 것에 강점이 있는 딥러닝 모형이다[16]. CNN은 다양한 구성으로 층을 쌓을 수 있는 블록(block)을 사용하는데, 이는 컨볼루션(convolution), 활성화(activation), 풀링(pooling)으로 이루어진다. Fig. 4는 블록의 구조를 보여주고 있다.

컨볼루션은 필터(filter)를 사용하여 이미지의 특징을 추출하는 작업이다. 필터가 자리를 옮기면서 이미지의 픽셀과 연산을 하는데, 이를 스트라이드(stride)라 부른다. 활성화 작업은 컨볼루션에서 추출된 출력값을 조절하여 비선형으로 변환하는 것이다. 이때 시그모이드(sigmoid)나 ReLU(ReLU : Rectified Linear Unit, 이하 ReLU) 등 비선형함수가 쓰이는데, 비선형함수는 선형함수와 달리 층을 쌓을 수 있어서 모형의 성능을 높이는 중요한 역할을 한다. 풀링은 입력 행렬에서 요소의 최대값을 반환하는 작업이다. 풀링은 크게 두 가지 역할을 하는데 하나는 차원을 축소하는 것이고, 다른 하나는 최대값을 취함으로써 작은 잡음(noise)을 제거하는 역할이다. 이 블록들을 층별로 쌓음으로써 작은 요소부터 큰 영역까지의 이미지를 표현할 수 있다.

마지막 빌딩 블록의 출력은 평면화 층(flatten layer)을 통해 1차원 벡터로 변환되고, 이 값들은 완전연결계 층(FC : Fully Connected layer, 이하 FC)으로 연결되며 소프트맥스(softmax)를 통해 최종 예측값으로 도출된다.

CNN을 이용한 주가 예측 연구는 LSTM(long short term memory) 등을 이용한 시계열 예측 연구와 함께 국내외에서 활발히 연구되고 있다[17]. Jiang et al.(2020)

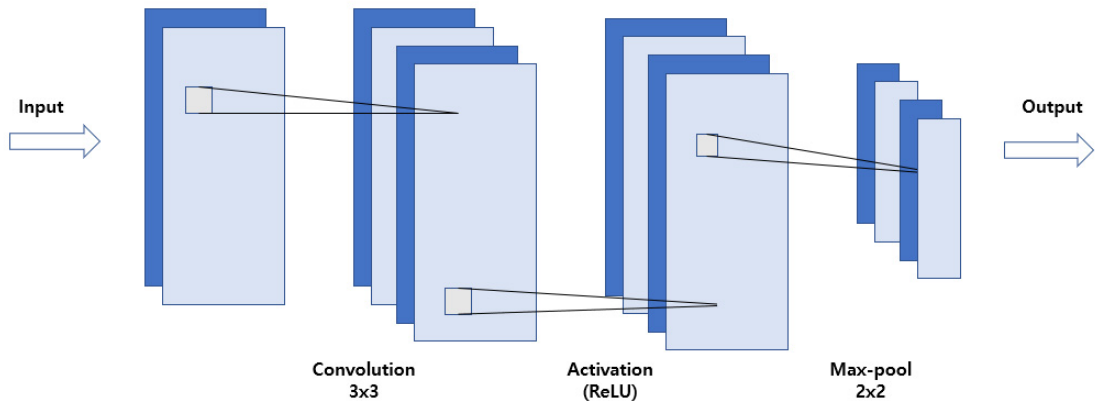


Fig. 4. Diagram of a building block

은 CNN과 이미지를 이용한 예측이 일반적 문헌에서 사용하는 추세 신호보다 정확하고, 수익성이 있으며 견고하다고 주장하였다. 또한, 일간 자료로 학습한 모형은 월간 자료를 예측하는 데 유용하며, 미국 주식에서 학습한 패턴은 국제 시장에서도 예측력이 높음을 보여주었다[18].

Sim et al.(2020)은 S&P500 지수에 9개의 기술적 지표(technical indicator)를 적용 후 이미지로 변환시켜 학습한 결과 CNN이 ANN(artificial neural network)이나 SVM(support vector machine)보다 정확도가 높음을 밝혔다. 그리고 이 연구에서는 CNN에 사용되는 하이퍼파라미터(hyperparameter)를 최적화하여 변수가 결핍함에 미치는 영향을 비교하였다[19].

Sezer et al.(2018)은 CNN을 통해 매수, 매도 신호를 예측하고 알고리즘 매매에 적용하는 연구를 하였다. 시뮬레이션 결과 매수 후 보유하는 전략보다 제안모형이 더 나은 수익성을 보임을 밝혔다[20].

Song(2018)은 국내 주가 자료를 이용하여 CNN으로 주가 매매 빈도 패턴과 예측 시점을 연구하였다[21]. KOSDAQ 상위 150종목을 대상으로 5분, 10분, 30분, 60분 단위의 차트를 만들어 주가의 등락을 예측하였다. 예측 결과 10분 단위 차트로 장 마지막 10분 전에 종가의 등락을 예측하는 경우 76%의 정확도를 보였다. 자료의 분석 기간은 장 마감 시점이 15시였던 2016년 8월 이전으로, 장 마감 시점이 15시 30분으로 바뀐 현재는 결과가 다를 수 있다는 한계점이 있다.

Lee and Ahn(2018)은 CNN을 통해 KOSPI200 주가지수의 등락을 예측하였다[22]. 2009년 1월부터 2017년 2월까지의 데이터를 이용해 대표적인 12개의 기술적 지표를 만들어 CNN을 통해 학습한 결과, CNN 모

형의 예측 성과가 비교모형인 로지스틱 회귀모형, 단일 계층 신경망, 그리고 SVM보다 높게 나타났다.

### 3. 자료 및 연구 모형

#### 3.1 자료

본 연구에서는 KOSPI200을 구성하는 전 종목의 1분 간격 자료와 일별 자료, KOPI200, S&P500, NASDAQ의 일별 자료를 이용하였다. Table 1과 Fig. 5는 학습에 사용된 자료 설명과 이미지 예시이다.

Fig. 5에서 ①은 한국엔컴퍼니(000240)의 2020년 6월 5일 1분 단위 시가(O), 고가(H), 저가(L), 종가(C)를 캔들스틱 차트로 나타낸 것이고, ②는 같은 기간의 1분 거래량(V)이다. ③은 해당 종목의 일별 종가를 이용하여 이동평균(MA)을 산출해 표시한 것이다. ④,⑤,⑥도 각각 KOSPI, S&P500, NASDAQ의 이동평균선을 보인 것이다. ⑤와 ⑥은 시차를 고려하여 전일까지의 자료를 이용하였다.

Table 1. Explanation of training image

	Name	Timeframe	Date	Data
①	Individual stock	Minute	T	O, H, L, C
②	Individual stock	Minute	T	V
③	Individual stock	Day	T	MA(3,5,10,20)
④	KOSPI	Day	T	MA(3,5,10)
⑤	S&P500	Day	T-1	MA(3,5,10)
⑥	NASDAQ	Day	T-1	MA(3,5,10)

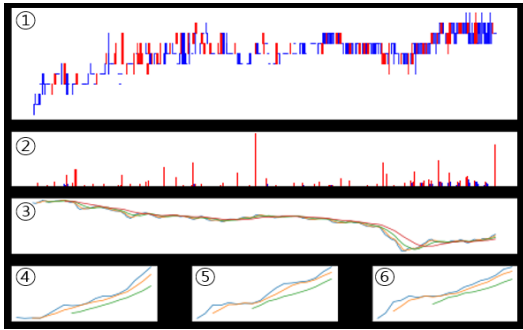


Fig. 5. Example of candlestick chart for training

주가 갭 상승과 하락을 예측하는 이진 분류(binary classification)를 위해 (2)와 같이 이미지에 0과 1을 라벨링 하였다.

$$Label\ for\ each\ image = \begin{cases} 0, & if\ gap > 0, \\ 1, & if\ gap \leq 0 \end{cases} \quad (2)$$

### 3.2 기초통계량

Table 2는 KOSPI200 전 종목의 주가 갭 수익률에 대한 기초통계량이다.

전체 표본 245,492개의 자료 중 갭 상승은 109,845개이고 갭 하락은 135,647개이다. 평균 수익률(average, avg)은 전체에서 0.0458%가 나왔고, 갭 상승과 갭 하락은 각각 0.8551%, -0.6096%를 보였다. 표준편차(standard deviation, stdev)의 경우 갭 상승은 1.0273%, 갭 하락은 0.9421%로 갭 상승에서 변동성이 더 컸는데, 이는 갭 하락에서 수익률 0을 포함하였기 때문으로 보인다. 갭 수익률의 t-통계량을 확인한 결과 18.57로 유의한 양(+)의 값이 나왔는데, 이는 표본 기간에 주가 갭 현상이 나타났음을 보여준다.

왜도(skewness, skew)는 0.88로 양으로 치우쳐있고, 첨도(kurtosis, kurt)는 57.11로 뾰족한 모양의 분포를 보인다. Fig. 6은 주가 갭 수익률의 히스토그램이다.

Table 2. Summary statistics of price gap return

	Total	Gap>0	Gap<=0
Count	245,492	109,845	135,647
Avg(%)	0.0458	0.8551	-0.6096
Stdev(%)	1.2219	1.0273	0.9421
t-statistic	18.57		
Skew	0.88		
Kurt	57.11		

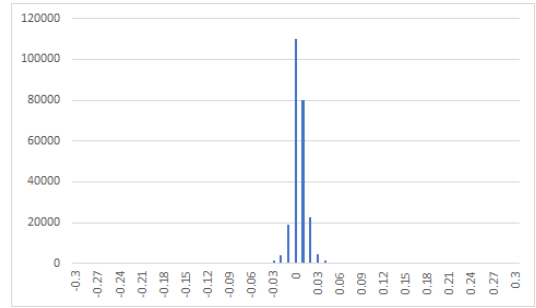


Fig. 6. Histogram of gap return

### 3.3 연구 모형

캔들스틱 차트 이미지를 이용하여 주가 갭을 예측하기 위해 이미지 파일을 Table 3과 같이 학습, 검증, 테스트로 데이터 세트(data set)를 구분하였다. 이때 갭 하락과 갭 상승의 이미지 수를 똑같이 맞추기 위해 갭 하락 이미지 수를 갭 상승의 이미지 수에 맞게 무작위 추출(random sampling)하였다[23]. 표본 기간은 2017년 1월부터 2021년 12월 30일까지이다. 이미지 입력 시 크기는 256x256, 채널(channel)은 3(RGB)으로 설정하였다.

Table 3. Training/validation/test data set

	Total	Training	Validation	Test
Count	219,690	153,782	32,954	32,954
Ratio	1	0.7	0.15	0.15

Fig. 7은 학습에 이용한 CNN 모형의 도표이다. 모형의 설계는 주가 이미지를 통해 추세를 예측하는 Jiang et al.(2020)의 연구 모형을 활용하였다[18]. 구체적으로는 Simonyan and Zisserman(2014)이 VGG에서 이용한 것처럼 첫 번째 층에 64개의 필터를 사용하였고[24], 층이 깊어질수록 모형이 복잡해지므로 컨볼루션 후 필터를 2배씩 늘려주었다[25].

활성화 함수는 Leaky ReLU(LReLU)를 이용하였는데, 이 함수는 ReLU와 달리 입력값이 음수일 때도 약간의 기울기를 부여하여 정보가 손실되지 않는다는 장점이 있다[26]. 식 (3)은 LReLU의 공식이고, Fig. 8은 ReLU와 LReLU의 차이를 도식화한 것이다.

$$LReLU(x) = \begin{cases} x, & if\ x \geq 0 \\ kx, & if\ x < 0 \end{cases} \quad where\ k = 0.01 \quad (3)$$

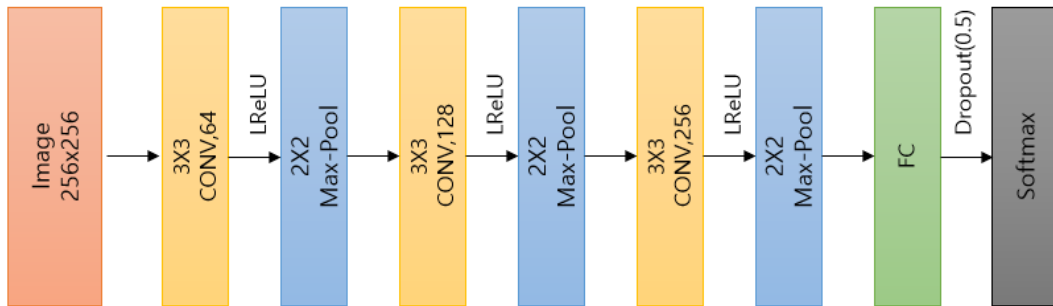


Fig. 7. Diagram of CNN model

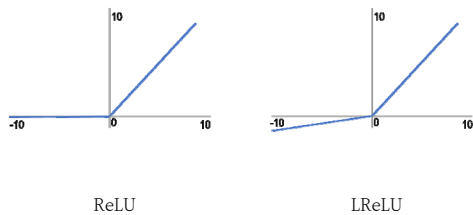


Fig. 8. A comparison between ReLU and LReLU

완결연결계층에서는 드롭아웃(dropout)을 이용하였다. 드롭아웃은 노드(node)를 특정 확률로 생략하는 방법으로 모형의 안정성을 높인다[27].

Table 4는 모형 설계에 사용된 변수를 보여주고 있다. 손실함수는 범주형 교차 엔트로피(categorical cross entropy)를 이용하였는데, 출력층에서 소프트맥스를 통해 꺾 상승과 하락의 확률을 출력하기 위함이다. 최적화 함수(optimizer)는 SGD(SGD : Stochastic Gradient Descent, 이하 SGD)를 사용하였다[13].

Table 4. Parameter settings

Parameter	Setting
Loss function	Categorical cross-entropy
Optimizer	SGD
Learning rate	0.01
Batch size	32
Dropout	0.5
Epoch	100

## 4. 분석 결과

### 4.1 모형 학습 및 검증 결과

제안한 연구 모형을 이용하여 학습(training) 및 검증(validation) 작업을 진행하였다. Fig. 9는 반복 횟수에 따른 정확도(accuracy)이고 Fig. 10은 반복 횟수에 따른 손실(loss)이다. 학습이 진행될수록 정확도는 올라가고 손실은 낮아지는 경향을 보였다. 검증에서의 정확도, 손실 결과도 학습과 비슷한 결과를 나타내 모형이 안정적인 모습을 보여주었다.

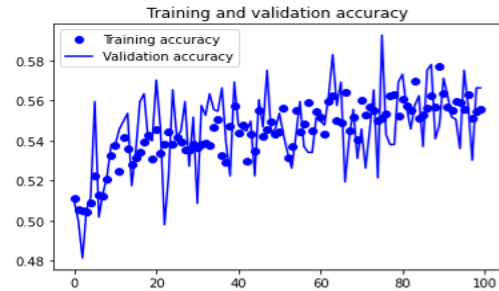


Fig. 9. Training and validation accuracy

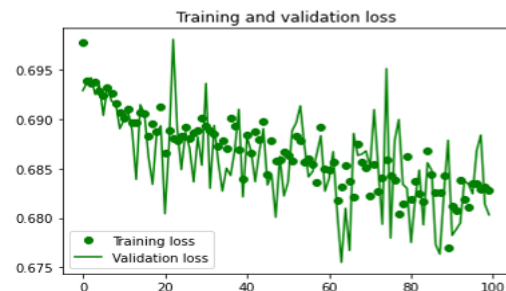


Fig. 10. Training and validation loss

최종 모형은 손실이 가장 낮았던 64번째 모형으로 선정하였다. 최종 모형의 정확도와 손실은 Table 5에 표시하였다.



Table 5. Accuracy and loss of selected model

	Training	Validation
Accuracy	0.5623	0.5830
Loss	0.6832	0.6755

### 4.2 모형 예측 결과

Table 6은 선정된 모형으로 테스트 데이터를 예측한 결과를 오 분류표(confusion matrix)로 나타낸 것이고, Table 7은 모형의 성능을 평가하기 위해 평가지표를 표로 나타낸 것이다. 평가지표는 (4)와 같이 계산된다.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (4)$$

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

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

$$F1 = \frac{2 * Precision * recall}{Precision + Recall}$$

where

TP is true positive

FP is false positive

FN is false negative

TN is true negative

Table 6. Confusion matrix of prediction

		Predict	
		Up	Down
Actual	Up	6,475	10,002
	Down	4,764	11,713

Table 7. Evaluation metrics of result

	Precision	Recall	F1-score
Up	0.58	0.39	0.47
Down	0.54	0.71	0.61
Accuracy			0.55

모형평가 결과 정확도(accuracy)는 0.55로 50% 이상의 예측률을 보였다.

정밀도(precision)에서 갭 상승은 0.58, 갭 하락은 0.54가 나왔다. 모형이 갭 상승으로 분류한 이미지 중 실제 갭 상승이었던 경우는 58%였고, 갭 하락으로 분류한 이미지 중 실제 갭 하락일 경우는 54%라는 뜻이다.

재현율(recall) 측면에서는 실제 갭 상승한 이미지 중 갭 상승으로 분류한 경우가 39%이고 실제 갭 하락한 이미지 중 갭 하락으로 분류한 경우가 71%였다. 전체적으로 갭 하락에 편향을 보였는데, 갭 하락 이미지에 수익률 0인 경우가 포함되어 있기 때문이라고 판단된다.

정밀도와 재현율은 상충관계(trade off)를 보이며, 투

자 결정에서는 재현율보다 정밀도가 더 중요한 평가지표로 판단된다. 재현율이 낮은 경우에는 단순히 투자 기회를 놓치는 것이지만, 정밀도가 낮은 경우는 자산 손실과 관련되기 때문이다.

F1점수는 각각 0.47, 0.61로 갭 하락 예측에서 더 좋은 성능을 보였다.

## 5. 투자전략 활용

### 5.1 표본 외 자료 예측 분석

학습, 검증, 테스트에 사용하지 않은 표본 외 자료를 모형을 통해 예측하여 데이터를 분석하였다. 분석 기간은 2022년 1월~2022년 8월 31일이다.

Fig. 11은 제안모형이 예측한 갭 상승 종목의 수와 갭 하락 종목의 수를 도식화한 것이다. 해당 기간 종목들의 하락을 우세하게 예측한 가운데, 지수가 하락추세일 때는 상승을 예측한 종목들이 많아지는 경향을 보였다.

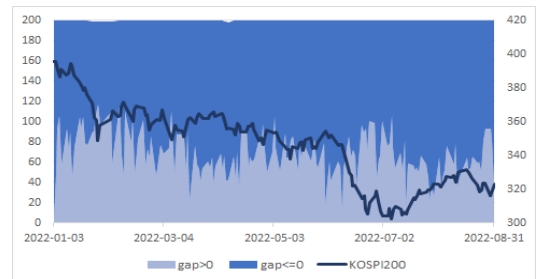


Fig. 11. The number of gap up and down

Table 8은 예측 결과값을 10분위로 나누는 후 그룹별 주가 갭 수익률의 일별 평균과 표준편차를 산출한 것이다. 1분위에서 평균 수익률은 0.1078%로 유의한 양의 수익률이 나타났으며, 분위가 높아질수록 음의 수익률이 높아지는 경향을 보였다.

Table 8. Summary statistics by decile

Decile	Avg(%)	Stdev(%)	t-statistic
1	0.1028	1.2788	4.7951
2	-0.0199	1.2750	-0.9307
3	-0.0703	1.1765	-3.5663
4	-0.0742	1.1990	-3.6931
5	-0.0663	1.1511	-3.4347
6	-0.0834	1.1195	-4.4451
7	-0.0728	1.1918	-3.6437
8	-0.0851	1.0503	-4.8323
9	-0.0834	1.0034	-4.9593
10	-0.1244	0.9578	-7.7513

### 5.2 L-S 전략의 수익성 분석

이번 장에서는 갭 상승 예측률 상위 1분위는 매수, 10분위는 매도하는 룡숏 전략을 시뮬레이션하여 수익성을 분석하였다[28]. 벤치마크(benchmark)로는 KOSPI200을 추종하는 상장지수펀드(exchange traded fund)인 KODEX200을 매일 증가에 매수하고 다음 날 시가에 매도하는 오버나이트전략(overnight strategy, OS)과 KODEX200을 매수 후 보유하는 전략(buy and hold, BH)으로 선정하였다. 위험을 고려하는 성과지표로는 위험조정성과(risk-adjusted return)인 수정 샤프비율(modified sharpe ratio, MSR)을 이용하였으며, MSR 계산 식은 (5)와 같다.

$$MSR = \begin{cases} \frac{avg*252}{stdev*\sqrt{(252)}}, & \text{if } avg > 0 \\ \frac{avg*252*stdev*\sqrt{(252)}}{stdev*\sqrt{(252)}}, & \text{if } avg \leq 0 \end{cases} \quad (5)$$

수정 샤프비율을 이용하는 이유는 수익률이 음수일 경우, 변동성이 작은 전략이 성과 상위에 있어야 하기 때문이다. 그리고 승률(win rate, WR)도 산출하여 보여주었다.

Table 9과 Fig. 12는 투자전략의 성과를 보여주고 있다.

Table 9. Performance of L-S strategy

Strategy	Avg(%)	Stdev(%)	MSR	WR(%)
L-S	0.2286	0.4275	8.4893	75.00
OS	-0.0291	0.7684	-0.0089	51.83
BH	-0.1119	1.1599	-0.0519	48.17

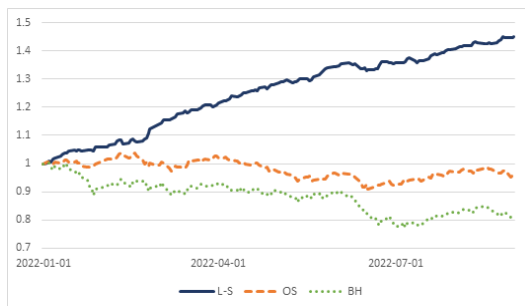


Fig. 12. Equity curve of L-S strategy

투자전략의 일별 평균 수익률은 0.2286%로, 동 기간 OS(-0.0291%)와 BH(-0.1119%)보다 우수한 성과를 보여주었다. L-S의 수정 샤프비율은 8.4893으로 거래비용(transaction cost)을 고려하지 않았을 때 벤치마크보다 우수한 성과가 나타났다.

### 5.3 거래비용을 고려한 L-S 전략 수익성 분석

Table 10과 Fig. 13은 거래비용을 고려한 L-S 전략의 성과이다.

Table 10. Performance of L-S strategy including cost

Strategy	Avg(%)	Stdev(%)	MSR	WR(%)
L-S(0.1%)	0.1286	0.4275	4.7764	65.85
L-S(0.2%)	0.0286	0.4275	1.0634	53.66
L-S(0.3%)	-0.0714	0.4275	-0.0122	43.29
OS	-0.0291	0.7684	-0.0089	51.83
BH	-0.1119	1.1599	-0.0519	48.17

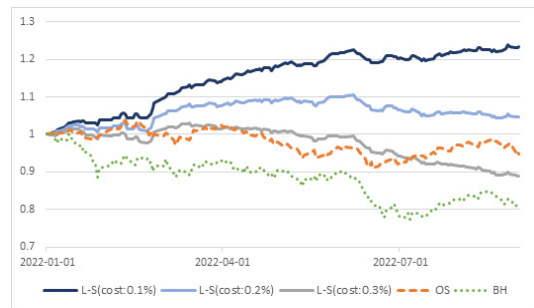


Fig. 13. Equity curve of L-S strategy including cost

거래비용이 0.1%일 때는 수정 샤프비율이 4.78로 벤치마크보다 우수하다. 그러나 증권거래세(0.23%)와 수수료 및 슬리피지(slippage)를 고려하여 0.3%의 거래비용을 고려하는 경우, 비교모형인 BH의 성과는 여전히 높지만, OS의 성과보다는 낮게 나타났다. 증권거래세가 2023년~2024년에는 0.2%, 2025년 이후에는 0.15%로 하향되는 것을 고려하면, 향후 실무적으로 충분히 활용될 수 있을 것으로 보인다.

## 6. 결론 및 연구의 한계점

본 연구는 KOSPI200 개별종목의 일중, 일별 그리고 해외 주가지수 정보를 담은 이미지를 CNN으로 학습한 후, 전일 증가와 당일 시가 사이의 수익률인 주가 갭을 예측하였다. 2017년 1월 ~ 2021년 12월 31일 자료를 7:1.5:1.5로 학습, 검증, 테스트 데이터 세트로 나누어 모형을 생성하였고, 모형을 통해 2022년 1월~2022년 8월 31일 자료의 갭 상승 확률을 예측하여 룡숏 투자전략을 시뮬레이션하였다. 실증분석 결과는 다음과 같다.

첫째, 학습된 CNN 모형은 갭 상승과 하락을 55%의



정확도로 맞추었다. 시장이 효율적이라면 다음날의 주가가 오를지 내릴지 예측하는 것은 불가능하다. 따라서 주가 등락 예측에 있어 55%의 정확도는 의미 있는 결과라 할 수 있다.

둘째, 갭 상승과 갭 하락을 나누어 분석한 결과, 갭 상승에서는 58%의 정밀도와 39%의 재현율을 갭 하락에서는 54%의 정밀도와 71%의 재현율을 보였다. 투자 결정에 관해서는 상승으로 분류한 주식이 실제 상승하는 확률인 정밀도가 실제 상승한 주식을 상승주식으로 분류하는 재현율보다 더 중요한 평가지표이므로, 정밀도가 높은 결과는 모형의 우수성을 보여주고 있다.

셋째, 표본 외 데이터를 이용한 시뮬레이션 결과 CNN 모형은 거래비용을 포함하고도 벤치마크인 KODEX200을 증가에 사서 시가에 매도하는 전략과 매수 후 보유하는 전략보다 우수한 성과를 보여주었다. 국내 증권시장이 선진화됨에 따라 증권거래세가 하향되고 있는 경향을 고려하면 추후 실무적으로 활용도가 높을 것으로 판단된다.

본 연구는 이미지에 개별주식의 일중 데이터뿐만 아니라 일간 이동평균선과 S&P500, NASDAQ의 이동평균선 정보를 담아 CNN 모형에 학습시킨 점에서 기존 연구와 차별점이 있지만, 다음과 같은 한계점도 있다.

첫째, 주가 갭 수익률에 영향을 미치는 다른 정보들을 넣지 않은 것은 한계점이다. 예를 들어 Kwon et al.(2015)에 따르면 국내 주식시장에서 야간수익률(overnight return)은 외국인 투자자가 주도한다고 밝혔는데[29], 일별 외국인 투자자 수급 추이 등 다른 정보를 포함하여 연구해볼 필요가 있다.

둘째, 시뮬레이션 방식을 고려해볼 필요가 있다. 본 연구의 방식에서는 2017년부터 2021년 데이터로 모형을 생성하고 2022년 데이터로 전진 분석(walk forward testing)을 진행하였는데, 이 방식은 학습된 기간에 발생한 이벤트에 의해 편향이 발생할 수 있는 단점을 가지고 있다. Lopez de Prado(2018)가 제안한 조합적 제거 교차 검증 기법(combinatorial purged cross validation)을 이용한다면 모형의 안정성과 신뢰성을 더 확실히 검증할 수 있을 것이다[30].

## References

- [1] C. Lee, "Artificial intelligence-based stock investment strategy : Buy-and-sell decision", *Korean management science review*, Vol.6, No.1, pp.71-84, 1989.
- [2] W. Lee, "A deep learning analysis of the KOSPI's directions", *Journal of the Korean Data and Information Science Society*, Vol.28, No.2, pp.287-295, 2017. DOI : <http://dx.doi.org/10.7465/jkdi.2017.28.2.287>
- [3] H. Choi, S. Kim, S. Park, "Analysis of trading performance on intelligent trading system for directional trading", *Journal of Intelligence and Information Systems*, Vol.17, No.3, pp.187-201, 2011. DOI: <http://dx.doi.org/10.13088/IIIS.2011.17.3.187>
- [4] Y. Kim, N. Kim, S. Jeong, "Stock-index invest model using news big data opinion mining", *Journal of Intelligence and Information Systems*, Vol.18, No.2, pp.143-156, 2012. DOI : <https://doi.org/10.13088/IIIS.2012.18.2.143>
- [5] S. Kim, "Forecasting COVID-19 pandemic stock prices using portal search intensity and deep learning", *Journal of Digital Contents Society*, Vol.23, No.2, pp.343-350, 2022. DOI : <http://dx.doi.org/10.9728/dcs.2022.23.2.343>
- [6] K. Lee, I. Yang, G. Jo, "Expert system for predicting the stock market timing using candlesticks chart", *Journal of Intelligence and Information Systems*, Vol.3, No.2, pp.57-70, 1997.
- [7] J. Lee, "A stock trading system based on moving average patterns and turning point matrix", *KIISE Transactions on Computing Practices*, Vol.18, No.7, pp528-532, 2012.
- [8] H. Yang, S. Kim, H. Choi, "Development of a stock trading system using M&W wave patterns and genetic algorithms", *Journal of Intelligence and Information Systems*, Vol.25, No.1, pp.63-83, 2019. DOI : <http://dx.doi.org/10.13088/jiis.2019.25.1.063>
- [9] F. Cross, "The behavior of stock prices on Fridays and Mondays", *Financial Analysts Journal*, Vol.29, No.6, pp.67-69, 1973. DOI : <https://doi.org/10.2469/FAJ.V29.N6.67>
- [10] K. R. French, "Stock returns and the weekend effect", *Journal of Financial Economics*, Vol.8, No.1, pp.55-69, 1980. DOI:[https://doi.org/10.1016/0304-405X\(80\)90021-5](https://doi.org/10.1016/0304-405X(80)90021-5)
- [11] D. B. Keim, R. F. Stambaugh, "A further investigation of the weekend effect in stock returns", *Journal of Finance*, Vol.39, No.3, pp.819-835, 1984. DOI : <https://doi.org/10.1111/j.1540-6261.1984.tb03675.x>
- [12] M. R. Gibbons, P. Hess, "Day of the week effects and asset returns", *Journal of Business*, Vol.54, No.4, pp.579-596, 1981.
- [13] J. L. Grant, A. Wolf, S. Yu, "Intraday price reversals in the us stock index futures market: A 15-year study", *Journal of Banking & Finance*, Vol.29, No.5, pp.1311-1327, 2005. DOI: <https://doi.org/10.1016/j.jbankfin.2004.04.006>
- [14] H. Ham, D. Ryu, and R. I. Webb, "The effects of overnight events on daytime trading sessions", *International Review of Financial Analysis*, Vol.83, 102228, pp.1-14, 2022. DOI : <https://doi.org/10.1016/j.irfa.2022.102228>

- [15] S. Kim, "Information contents in overnight puzzle: Is puzzle strategy profitable?", *The Korea Contents Association*, Vol.22, No.11, pp.537-547, 2022.  
DOI : <https://doi.org/10.5392/JKCA.2022.22.11.537>
- [16] Y. LeCun, Y. Bengio, G. Hinton, "Deep learning", *Nature*, Vol.521, No.7553, pp.436-444, 2015.  
DOI : <https://doi.org/10.1038/nature14539>
- [17] J. Shah, D. Vaidya, M. Shah, "A comprehensive review on multiple hybrid deep learning approaches for stock prediction", *Intelligent Systems with Applications*, Vol.16, pp.2001111, 2022.  
DOI : <https://doi.org/10.1016/j.iswa.2022.200111>
- [18] J. Jiang, B. T. Kelly, D. Xiu, "(Re-)Imag(in)ing price trends". *Chicago Booth Research Paper*, No. 21-01, 2020.  
DOI : <http://dx.doi.org/10.2139/ssrn.3756587>
- [19] H. Sim, H. Kim, J. Ahn, "Is deep learning for image recognition applicable to stock market prediction?", *Complexity*, Vol.2019, pp.1-10, 2019.  
DOI : <https://doi.org/10.1155/2019/4324878>
- [20] O. B. Sezer, A. M. Ozbayoglu, "Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach.", *Applied Soft Computing*, Vol.70, pp.525-538., 2018.  
DOI : <https://doi.org/10.1016/j.asoc.2018.04.024>
- [21] H. Song, S. Lee, "A study on the optimal trading frequency pattern and forecasting timing in real time stock trading using deep learning: Focused on KOSDAQ", *The Journal of Information Systems*, Vol.27, No.3, pp.123-140, 2018.  
DOI : <https://doi.org/10.5859/KAIS.2018.27.3.123>
- [22] M. Lee, H. Ahn, "A time series graph based convolutional neural network model for effective input variable pattern learning : Application to the prediction of stock market", *Journal of Intelligence and Information Systems*, Vol.24, No.1, pp.167-181, 2018.  
DOI : <https://doi.org/10.13088/JIIS.2018.24.1.167>
- [23] C. Lim, "A study on stock trend determination in stock trend prediction", *Journal of the Korea Society of Computer and Information*, Vol.25, No.12, pp.35-44, 2020.  
DOI : <https://doi.org/10.9708/JKSCI.2020.25.12.035>
- [24] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition", *Computer Science - Computer Vision and Pattern Recognition*, working paper, 2014.  
DOI : <https://doi.org/10.48550/arXiv.1409.1556>
- [25] M. D. Zeiler, R. Fergus, "Visualizing and understanding convolutional networks", *Computer Science - Computer Vision and Pattern Recognition*, working paper, 2013.  
DOI : <https://doi.org/10.48550/arXiv.1311.2901>
- [26] A. L. Maas, A. Y. Hannun, A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models", *Proceedings of the 30th International Conference on Machine Learning*, Vol.28, pp.3, 2013.
- [27] P. Baldi, P. J. Sadowski, "Understanding dropout", *Advances in neural information processing systems*, Vol.26, 2013.
- [28] N. Jegadeesh, S. Titman, "Returns to buying winners and selling losers: Implications for stock market efficiency", *The Journal of Finance*, Vol.48, No.1, pp.65-91, 1993.  
DOI: <http://dx.doi.org/10.1111/j.1540-6261.1993.tb04702.x>
- [29] E. Kwon, Y. Eom, W. Jang, and J. Hahn, "Who overreacts to overnight news?: Empirical evidence from the Korean stock market", *Asia-Pacific Journal of Financial Studies*, Vol.44, pp.298-321, 2015.  
DOI: <https://doi.org/10.1111/aifs.12090>
- [30] M. Lopez de Prado, *Advances in Financial Machine Learning*, John Wiley & Sons, 2018.

김 찬 수(Chansu Kim)

[정회원]



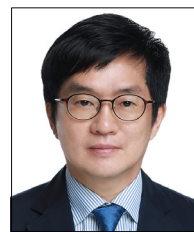
- 2016년 2월 : 인하대학교 토목공학과(공학사)
- 2018년 2월 : 국민대학교 비즈니스IT전문대학원(공학석사)
- 2016년 11월 ~ 현재 : 한국펀드평가 펀드연구원
- 2022년 3월 ~ 현재 : 국민대학교 비즈니스IT전문대학원 박사과정

<관심분야>

알고리즘트레이딩, 계량분석

최 흥 식(Heung Sik Choi)

[정회원]



- 1983년 2월 : 한양대학교 산업공학과(공학사)
- 1985년 2월 : KAIST 경영과학과(공학석사)
- 1995년 2월 : 로체스터대 컴퓨터정보시스템(경영학박사)
- 1995년 3월 ~ 현재 : 국민대학교 비즈니스IT전문대학원 교수

<관심분야>

트레이딩시스템, 증권투자연구

김 선 웅(Sun Woong Kim)

[정회원]



- 1981년 2월 : 서울대학교 경영학과 (경영학사)
- 1983년 2월 : KAIST 경영과학과 (공학석사)
- 1988년 2월 : KAIST 경영과학과 (공학박사)
- 2009년 3월 ~ 현재 : 국민대학교 비즈니스IT전문대학원 교수

〈관심분야〉

트레이딩시스템, 투자위험관리