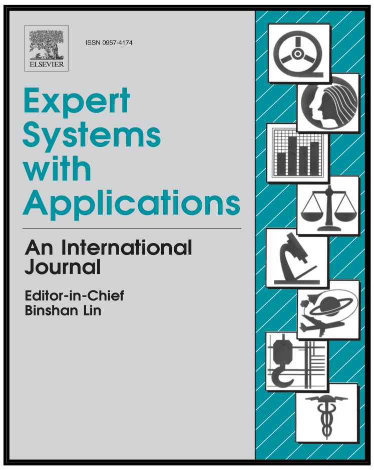
Accepted Manuscript



Bankruptcy prediction using imaged financial ratios and convolutional neural networks

Tadaaki Hosaka

PII: S0957-4174(18)30616-X
DOI: https://doi.org/10.1016/j.eswa.2018.09.039
Reference: ESWA 12227

To appear in: *Expert Systems With Applications*

Received date: 16 May 2018
Revised date: 20 July 2018
Accepted date: 17 September 2018

Please cite this article as: Tadaaki Hosaka, Bankruptcy prediction using imaged financial ratios and convolutional neural networks, *Expert Systems With Applications* (2018), doi: https://doi.org/10.1016/j.eswa.2018.09.039

Highlights

- Convolution networks can predict bankruptcy by inputting financial ratios as an image
- Predictive accuracy improves with correlated financial ratios placed in the vicinity
- Deeper network configuration improves predictive accuracy
- Creating artificial financial data does not ensure the same effect as using real data
- Convolution-network-based bankruptcy prediction outperforms traditional methods

Bankruptcy prediction using imaged financial ratios and convolutional neural networks

Tadaaki Hosaka[1]

1-11-2 Fujimi, Chiyoda City, Tokyo, Japan

Tokyo University of Science[1,1]

Abstract

Convolutional neural networks are being applied to identification problems in a variety of fields, and in some areas are showing higher discrimination accuracies than conventional methods. However, applications of convolutional neural networks to financial analyses have only been reported in a small number of studies on the prediction of stock price movements. The reason for this seems to be that convolutional neural networks are more suitable for application to images and less suitable for general numerical data including financial statements. Hence, in this research, an attempt is made to apply a convolutional neural network to the prediction of corporate bankruptcy, which in most cases is treated as a two-class classification problem. We use the financial statements (balance sheets and profit-and-loss statements) of 102 companies that have been delisted from the Japanese stock market due to de facto bankruptcy as well as the financial statements of 2062 currently listed companies over four financial periods. In our proposed method, a set of financial ratios are derived from the financial statements and represented as a grayscale image. The image generated by this process is utilized for training and testing a convolutional neural network. Moreover, the size of the dataset is increased using the weighted averages to create synthetic data points. A total of 7520 images for the bankrupt and continuing enterprises classes are used for training the convolutional neu-

Email address: `hosaka-t@rs.tus.ac.jp` (Tokyo University of Science)

Preprint submitted to Elsevier *September 18, 2018*

ral network based on GoogLeNet. Bankruptcy predictions through the trained network are shown to have a higher performance compared to methods using decision trees, linear discriminant analysis, support vector machines, multi-layer perceptron, AdaBoost, or Altman's Z''-score.

Keywords: Deep learning; business failure; financial statement; imaging

1. Introduction

An accurate prediction of the future performance of an enterprise is important for investors to generate profits in securities trading. One typical example of such a prediction is the probability of a company going bankrupt. When companies fall into bankruptcy, there is a high possibility that they will be delisted, which has a major impact on investors. Hence, in stock exchanges, companies that are at high risk of being delisted are designated as supervised stocks. It is thus extremely beneficial for investors to recognize the signs of bankruptcy at an early stage.

1.1. Previous research on bankruptcy prediction using machine learning

Much research on bankruptcy prediction has been performed using pattern recognition or machine learning. In 1968, Altman (1968) investigated the performance of linear discriminant analysis for 22 financial ratios that he selected based on his knowledge of accounting and constructed a prediction model using five of the financial ratios. Since then, non-linear prediction models using logistic regression (Ohlson, 1980), neural networks (Odom and Sharda, 1990; Brédart, 2014), support vector machines (SVM) (Shin, Lee, and Kim, 2005; Li and Sun, 2009), and AdaBoost (Alfalo, García, Gámez, and Elizondo, 2008; Ramakrishnan, Mirzaei, and Bekri, 2015) have been proposed. Besides the genetic algorithm (Back, Laitinen, and Sere, 1996; Gordini, 2014) which is usually utilized in optimization problem, and case-based reasoning (Bryant, 1998; Sartori, Mazzucchelli, and Gregorio, 2016) have been applied to acquiring the decision rule for bankruptcy. The existing research on bankruptcy prediction manually defines several (in most cases, around three to five) financial indicators which are useful for constructing prediction models. One potential problem with existing methods is that financial indicators which are potentially suitable for identification may be incorrectly passed over, as they have a complex non-linear relationship with the probability of bankruptcy.

Shirata (1998) has applied statistical methods to the selection of financial indicators to avoid this problem. She used the Classification and Regression Tree (CART) algorithm to select four financial indicators which are effective for bankruptcy prediction, out of the 61 financial indicators deemed appropriate from an accounting viewpoint, and derived linear discriminant functions using these four variables. However, the selection of the financial indicators and the construction of the prediction model via machine learning are separate processes; hence, there remains the problem that the optimality of the prediction accuracy for the method as a whole cannot be guaranteed.

To overcome this problem, Takata, Hosaka, and Ohnuma (2015, 2017) utilized AdaBoost and an advanced version, RealAdaBoost (Hosaka and Takata, 2016) to execute both processes within a single framework, in which their algorithms determine the effective financial ratios from a large number of candidates and construct a prediction model by weighted majority voting with the selected indicators. Other approaches have also been considered which do not use financial information at all; for example, Shirata and Sakagami (2008) predicted corporate bankruptcy by applying text mining techniques to financial statements or corporate investor relations (IR) documents.

1.2. Previous research using deep learning for financial analysis

Deep learning is gathering great attention in the field of machine learning and artificial intelligence, and has achieved great success, especially for image recognition (e.g. Krizhevsky, Sutskever, and Hinton, 2012; Lin, Chen, and Yan, 2013; Szegedy, Liu, Jia, Sermanet, Reed, Anguelov, Erhan, Vanhoucke, and Rabinovich, 2015), voice recognition (e.g. Mohamed, Dahl, and Hinton, 2012; Dahl, Yu, Deng, and Acero, 2012; Graves, Mohamed, and Hinton, 2013), and natural language processing (e.g. Yao, Zweig, Hwang, Shi, and Yu, 2013; Irsoy and Cardie, 2014; Kiros, Salakhutdinov, and Zemel, 2014). However, the number of applications of deep learning to financial analysis is extremely limited except for several reports on the prediction of stock price fluctuations (Yoshihara, Fujikawa, Seki, and Uehara, 2014; Aggarwal and Aggarwal, 2016; Persio and Honchar, 2017; Chong, Han, and Park, 2017; Bao, Yue, and Rao, 2017), where recurrent neural networks are often utilized for time series analysis.

An example of using deep learning in financial analysis other than for forecasting stock price fluctuations is provided by Yeh, Wang, and Tsai (2015), who predicted the probability of bankruptcy via a deep belief network (DBN) with five layers, where changes in stock price volatility are represented as a line graph (a binary image) and each pixel value (1 or 0) corresponds to a visible variable. They reported that the discrimination performance for identifying bankrupt and continuing companies is superior to the conventional SVM algorithm. Lee, Jang, and Park (2017) also estimated the gross sales amount, operating income, and net income by using a DBN. They focused on not only financial indicators from the preceding fiscal year but also the number of patent applications and the patent shares within the industry three years prior. These features are utilized as the visible variables for the learning of a DBN with six layers.

Convolutional neural networks (CNNs) have been producing drastic improvements in the performance of image recognition algorithms (Krizhevsky, Sutskever, and Hinton, 2012; Lin, Chen, and Yan, 2013; Szegedy, Liu, Jia, Sermanet, Reed, Anguelov, Erhan, Vanhoucke, and Rabinovich, 2015). However, there are only a few examples of CNNs being applied to financial analyses. The reason for this seems to be that CNNs are more suitable for application to images and less suitable for general numerical data including financial statements.

As an example of one of the few applications of a CNN to financial data, Ding, Zhang, Liu, and Duan (2015) used a CNN to predict whether a share price would rise or fall. They used texts describing finance-related events from the news over the previous month and vectorized them to use as input for their CNN. Convolutional operations on data for consecutive days were performed for feature extraction, and the subsequent maximum pooling layer leaves only the most influential variable. It was reported that there was an improvement in the estimation accuracy using a network containing four layers. Siripurapu (2015) also proposed expressing fluctuations between high and low prices of the S&P500 over a period of 30 minutes with images (typically line graphs) and utilizing them as input for a CNN to predict the price in the subsequent five minutes. However, a significant improvement in the prediction accuracy was not observed.

Based on the above background, the present research aims to propose an effective method for applying CNNs to bankruptcy prediction. Since CNNs are particularly effective for image analysis, financial numerical data need to first be converted before constructing a CNN. Hence, we propose a transformation for converting a set of financial ratios generated from a balance sheet and a profit-and-loss statement into an image. More specifically, each financial ratio corresponds to a particular pixel, and the brightness of the pixel is determined by the value of the corresponding financial ratio. Images created in this way represent a single enterprise for a particular fiscal year and serve as training data for the CNN. Also, the amount of data is increased through weighted averaging of financial statements spanning multiple fiscal years. Our CNNs based on GoogLeNet (Szegedy, Liu, Jia, Sermanet, Reed, Anguelov, Erhan, Vanhoucke, and Rabinovich, 2015) are trained for the two-class identification problem of whether the target company belongs to the bankrupt enterprises class or the continuing enterprises class. Application of our method to real-world data shows that our prediction performance is superior to that of conventional methods.

2. Data

In this research, enterprises delisted due to business failures in one of the Japanese stock markets (Tokyo Stock Exchange, Osaka Securities Exchange, former Nasdaq Japan Standard, former Hercules Standard, former Hercules Growth, or the former Jasdaq) between January 2002 and June 2016 are considered to be bankrupt companies. The reasons for delisting, which can be viewed as de facto bankruptcy, are 1) bankruptcy/rehabilitation/reorganization procedures, 2) excessive debt, 3) suspension of bank transactions, and 4) termination of business activities (excluding mergers). A total of 153 companies falling under these criteria are used in this research. The continuing companies in this research are 2450 firms listed on the Tokyo Stock Exchange First Section or Second Section as of June 2016.

Two types of financial statements, the consolidated balance sheets and profit-and-loss statements, of these companies are obtained from the Nikkei NEEDS Financial QUEST database for four periods preceding their delisting (bankrupt companies) or before June 2016 (continuing companies). There are 175 items included in a balance sheet and 88 items in a profit-and-loss statement. By handling the data for each year separately, four entries can be obtained from a single company. Assuming that listed companies rarely go bankrupt within a year or two of a deterioration in management, the four financial data from a bankrupt company are all treated as belonging to the "bankrupt enterprises class".

There are missing values in the financial statements. This is because 1) the Japanese accounting standards for the net assets section changed in 2006, 2) the notation for accounting items can differ depending on the industry, even if they have similar meanings, and 3) items that have zero value are missing. Hence, some missing values are supplemented when other related accounting items are available to aid the estimation. After this operation, enterprises that still have missing values for the following variables are excluded from this research.

Requisite accounting variables:

1. Current assets
2. Fixed assets
3. Current liabilities
4. Fixed liabilities
5. Net assets

6. Shareholders' equity
7. Retained earnings
8. Sales volume
9. Cost of sales
10. Gross profit
11. Operating profit
12. Ordinary profit
13. Net profit before taxes
14. Net profit

Missing values other than for the above variables are left unaltered. Financial institutions such as banks are also excluded from the scope of our analysis. There are a total of 102 companies in the bankrupt enterprises class and 2062 companies in the continuing enterprises class.

3. Proposed method

In this research, we generate as many financial ratios as possible from the financial statements of each company in each fiscal year and express the set of ratios as a single grayscale image. To achieve this, each financial ratio is made to correspond to a specific pixel position (x, y-coordinates) and the brightness value of that pixel is set based on the value of the corresponding financial ratio. The images generated with this process are then used as input to train the CNN based on GoogLeNet (Szegedy, Liu, Jia, Sermanet, Reed, Anguelov, Erhan, Vanhoucke, and Rabinovich, 2015) (a representative CNN proposed by Google). Details of the proposed method are described below.

3.1. Correspondence between financial ratios and pixels

CNNs perform particularly well in image recognition (Krizhevsky, Sutskever, and Hinton, 2012; Lin, Chen, and Yan, 2013; Szegedy, Liu, Jia, Sermanet, Reed, Anguelov, Erhan, Vanhoucke, and Rabinovich, 2015). Hence, this research proposes that all the financial ratios generated from the balance sheet and profit-and-loss statement (per company) be represented in a single image. For this, each financial ratio needs to be made to correspond to a specific pixel. The following two methods are considered:

Random: Randomly determine the correspondence between the financial ratios and pixel positions,

Correlated: Determine the correspondence between financial ratios and pixel positions so that highly correlated financial ratios are placed as close as possible to one another.

The idea behind the 'Correlated' method is that natural images for object recognition generally have extremely large correlations between neighboring pixels. On the other hand, even the 'Random' method can take into account the relationship between distant pixels by making the CNN multi-layered. The evaluation experiment will compare the results between the two methods.

The details of our procedure are now described. Carrying out the following procedure up to step (4a) corresponds to the 'Random' method, whereas completion of all the steps corresponds to the 'Correlated' method. Note that although expressions such as 'image coordinates' or 'pixel positions' are more appropriate, they are simply referred to as 'pixels' in this paper unless there is a risk of confusion.

In the 'Correlated' method, we utilize the financial ratios generated from 2062 continuing companies' data within one year of June 2012 (the oldest data in the four periods; referred to hereinafter as the 'reference dataset') to find as typical correlation values as possible for sound Japanese companies. Thus, companies in the bankrupt enterprises class are not involved in the correlation calculation because the financial situation tends to drastically change over a few years prior to bankruptcy, which cannot be regarded as typical.

Procedures:

(1) For each of the financial items, if the fraction of companies having missing

values out of all the 2062 companies in the reference dataset is greater than $p (0 \leq p \leq 1)$, then that financial item is excluded.

(2) The remaining financial items are used to obtain financial ratios for each of the companies included in the reference dataset. Any two items are chosen from the balance sheet and profit-and-loss statement (including the case of selecting one item from both financial statements) and their ratio is computed. Since there is no fundamental difference in switching the numerator and the denominator, either one is considered. In this research, the financial item with the smaller average value in the reference dataset is set as the denominator. When the numerator or the denominator is a missing value, the corresponding financial ratio is also treated as a missing value. Besides the financial ratio whose denominator is zero in any company is excluded from explanatory variables.

(3) The correlation coefficients are calculated for all the combinations of generated financial ratios. Enterprises with missing values for the financial ratios considered are not included in the computation. Therefore, when there are no common non-missing enterprises in the related two financial ratios, it is not possible to obtain the correlation coefficient. In such a case, those financial ratios are excluded. Also, for financial ratios whose standard deviations are approximately zero, the correlation coefficient with other financial ratios cannot be obtained, and hence the financial ratio is excluded.

(4) For the total number of financial ratios, N, an image of $\lceil \sqrt{N} \rceil \times \lceil \sqrt{N} \rceil$ size is considered, and the correspondence between the financial ratios and pixels is determined through the following Monte Carlo simulation. Here, $\lceil a \rceil$ denotes the smallest integer greater than a.

a) For the initial setting, each financial ratio is set to correspond to a pixel at random with no overlap. The energy (objective function) related to the correspondence between financial ratios and pixels is defined as

$$E = \sum_{(i,j)\in\mathcal{P}} |c[\mathcal{R}(i), \mathcal{R}(j)]| \; d(i,j), \tag{1}$$

$$d(i,j) = \{x(i) - x(j)\}^2 + \{y(i) - y(j)\}^2, \tag{2}$$

where i, j are the indices for pixels, and $x(i), y(i)$ represent the x and y coordinates of pixel i, respectively. $\mathcal{P}$ represents the set of all combinations of pixels. $\mathcal{R}(i)$ represents the financial ratio that corresponds to pixel i and $c[\mathcal{R}_1, \mathcal{R}_2]$ represents the correlation coefficient between financial ratios $\mathcal{R}_1$ and $\mathcal{R}_2$.

b) If the energy (1) can be reduced by switching the financial ratios which correspond to two randomly selected pixels, the interchange is implemented. When equation (1) does not decrease, nothing is done.

c) Repeat step b). However, when there is no switching of financial ratios $3N$ times in a row, the simulation is terminated.

Applying this method to the reference dataset, the correspondence between the financial ratios and pixels can be obtained. Hereinafter, the notation $\mathcal{R}(i)$ is used to represent the financial ratios corresponding to a pixel i in the obtained correspondence.

3.2. Increasing the number of data points via weighted averages

CNN learning generally requires a large amount of training data. In this research, the number of entries belonging to the bankrupt enterprises class is small, at only 408 (102 companies, 4 periods). Hence the amount of data available for both classes is increased using the weighted averages approach as in the interpolation and extrapolation. The purpose of this synthetic data generation is similar to processing in image recognition where the sample sizes are increased by geometric transformation such as horizontal/vertical inversion and translation.

The specifics of the weighted average method are described below. A weighted average over two arbitrarily selected periods for each company and each financial item is computed via the equation:

$$wz_1 + (1-w)z_2, \tag{3}$$

where z_1 and z_2 represent the values of the financial item for the two selected years and w is a weighting factor. A total of 15 values for the coefficient w,

$$-0.6 \quad -0.4 \quad -0.2 \quad 0.1 \quad 0.2 \quad 0.3 \quad 0.4 \quad 0.5 \quad 0.6 \quad 0.7 \quad 0.8 \quad 0.9 \quad 1.2 \quad 1.4 \quad 1.6,$$

are empirically defined. We do not use values $w < -0.6$ and $w > 1.6$ because the accuracy of extrapolation generally becomes worse as the generated synthetic point goes farther away from the actual data point. When either z_1 or z_2 is missing, the weighted average is also regarded as a missing values. There are six ways of selecting two arbitrary periods; hence, this operation generates 6×15 synthetic data for one enterprise. Consequently, we have 94 data in total from one enterprise including the original values for the four year period.

3.3. Determination of pixel brightness from corresponding financial ratios

The correspondence between the financial ratios and pixels derived in section 3.1 is used to describe the real data for the four periods, as well as the synthetic data obtained in section 3.2, as a grayscale image. Firstly, the financial ratios are computed for each data using the same method as in step (2) in section 3.1. Then, the brightness of pixel i is determined by

$$\frac{v[\mathcal{R}(i)] - m[\mathcal{R}(i)]}{\sigma[\mathcal{R}(i)]} \times 100 + 128, \tag{4}$$

where $v[\mathcal{R}(i)]$ represents the value of the financial ratio $\mathcal{R}(i)$ of the data concerned. In the above equation, $m[\mathcal{R}(i)]$ and $\sigma[\mathcal{R}(i)]$ represent the average and standard deviation, respectively, of the financial ratios $\mathcal{R}(i)$ of all 2062 data included in the reference dataset. However, companies which have missing values for particular financial ratios are not included in the calculation of averages or standard deviations for those ratios. When the value of equation (4) is greater than the upper limit for the brightness of 255 or less than the lower limit of 0, it is taken to be 255 or 0, respectively. If a financial ratio is greater than the average of the reference dataset, then the corresponding pixel is assigned a color closer to white, and if it is less than the average of the reference dataset, then it is assigned a color closer to black. When $v[\mathcal{R}(i)]$ is a missing value or the denominator of the ratio is zero, the brightness value of pixel i is set to 128.

3.4. Training dataset and test dataset

The way of assigning the obtained grayscale images to training and test datasets is described below. Although our basic approach follows the five-fold cross-validation, there are two points to be noted. Firstly, since the number of bankrupt companies is much less than that of continuing companies, as in many conventional methods, we determine the size of learning data based on bankrupt enterprises class. As a result, most of the continuing enterprises are not used for learning and are included only in test datasets. Secondly, since the artificial data are considered unsuitable for evaluation in our case, images generated from the synthetic financial data are used only for learning.

The detailed procedures are as follows:

- From the 102 companies belonging to the bankrupt enterprises class, 20 companies are randomly selected. This process is repeated to generate a total of five subsets, with no company belonging to more than one subset. These subsets are referred to as groups A-E. The remaining two companies do not belong to any of the groups.
- From the 2062 companies belonging to the continuing enterprises class, 20 companies are randomly selected. This process is repeated to generate a total of five subsets, with no company belonging to more than one subset. The generated subsets are added to one of the above-mentioned groups A-E. The remaining 1962 companies do not belong to any of the groups.
- For both the bankrupt and continuing companies belonging to each group, images generated from the real data over the four periods and from the synthetic data obtained in section 3.2 are used to form datasets A-E.

- Four of datasets A-E are selected as the training datasets. The test dataset consists of the images generated from the real financial data of the enterprises in the remaining dataset (20 companies in both classes) and from the enterprises not belonging to any of the groups (two bankrupt companies and 1962 continuing companies). Therefore, it should be noted that although images generated from the synthetic financial data are included in the training dataset, they are not included in the test dataset. There are five ways to select four of datasets A-E; hence, five pairs of training and test datasets are obtained in total.

For both classes, 7520 images (80 companies $\times$ (4 periods + 90 synthetic values)) are contained in one training dataset generated via the above process. On the other hand, each test dataset contains 88 images (22 companies $\times$ 4 periods) for the bankrupt enterprises and 7928 images (1982 companies $\times$ 4 periods) for the continuing enterprises. Note that the two bankrupt companies and 1962 continuing companies not belonging to any of the groups A-E are included in all test datasets.

3.5. Learning and evaluation based on GoogLeNet

The learning of the CNN parameters is performed by using the training datasets generated in the aforementioned way. The CNN in this research is based on GoogLeNet (Szegedy, Liu, Jia, Sermanet, Reed, Anguelov, Erhan, Vanhoucke, and Rabinovich, 2015) which was awarded first place in the 2014 ImageNet Large Scale Visual Recognition Competition (ILSVRC). Although CNNs had already been used before the current enthusiasm of deep learning in, for example, LeNet (Lecun, Bottou, Bengio, and Haffner, 1998), its effectiveness became widely known by the Alexnet (Krizhevsky, Sutskever, and Hinton, 2012) comprising five convolutional layers and three full-connected layers. Although Alexnet presented high recognition performance, its learning process is highly inefficient because the number of parameters is very large (about 60 million) due to the successive full-connected layers. GoogLeNet whose name was attached in honor of LeNet, comprises 27 layers from the input to the output layers, and this network configuration was proposed to save computational time while further improving the identification accuracy. Although the number of layers was greatly increased, the number of parameters was fewer than Alexnet (about seven million).

The factors that made GoogLeNet successful were a convolution of the 1×1-sized filter proposed by Lin, Chen, and Yan (2013) and the inception module. Assuming that the number of input channels from the previous layer is K, the 1×1 convolutional operation corresponds to a linear combination of K pixel values having the identical coordinates, and dimension reduction can be achieved by setting the number of output channels smaller than K. GoogLeNet greatly reduces the number of parameters by frequently using the 1×1 convolutional operation.

As shown in Fig. 1, the inception module copies the input coming from the previous layer, passes them through the four paths, concatenates the four consequent outputs, and finally send them to the next layer. For each path, the size of the filter is fixed, and convolution of multiple filters having different coefficient values is performed in parallel. Those outputs are represented as a multi-channel image. Thus, the number of parameters is expressed as filter size $\times$ the number of input channels $\times$ the number of output channels. Although the equivalent task could be conducted without branching the path by using as many 5×5-sized filters as the sum of the output channels for the four paths, the reduction of the number of parameters are achieved by preparing multiple routes and setting smaller-sized filters. The leftmost route of the inception module consisting of only 1×1 convolution, directly transfers the information from the previous layer to the next layer, which can be regarded as the connection bypassing the distant layers. The effectiveness of the by-path in CNNs then attracted the attention of researchers (He, Zhang, Ren, and Sun, 2016). While stacking the inception modules, GoogLeNet has only one full-connected layer as the output layer after the image size is sufficiently reduced, which significantly reduces computational cost.

Table 1 shows five kinds of network structures in this research. The network

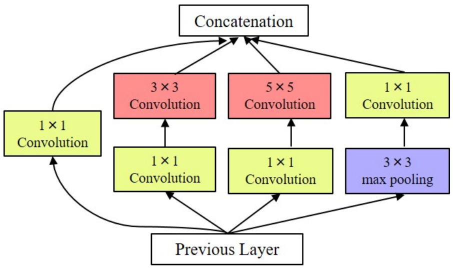


Figure 1: Inception module. Inputs coming from the previous layer are copied and passed through the four paths. An output from each route is finally concatenated to proceed to the next layer.

with 27 layers is the original GoogLeNet except that the number of output channels for each layer is roughly divided into quarters of the original configuration to reduce the computational load. Furthermore, we also evaluate the performance of smaller networks with 23, 17, 11, and 6 layers by deleting some constituents of the original network. The number of learning epochs and the mini-batch size are empirically set as 100 and 32, respectively.

The trained network is used to evaluate the identification rate of the test data. By repeating the experiments for five pairs of training and test datasets, the effectiveness of the proposed method is verified based on the average identification performance.

4. Evaluation experiments and discussion

4.1. Imaging of financial ratios

The results of expressing the financial ratios as grayscale images are now described. Setting values of the parameter p between 0 to 1 in increments of 0.1, we selected $p = 0.8$, which retained the greatest number of financial ratios, $N = 534$, in step (4) of section 3.1. Table 2 shows the 133 financial items in balance sheets and profit-and-loss statements which remained after step (1) of section 3.1. Hence, an image of size 24×24 pixels is generated. The energy for a random initial placement in step (4a) of section 3.1 was $E = 2.02 \times 10^6$

Table 1: Network configurations. The layer types in the table are ordered from the input layer on the uppermost row (Convolution) to the output layer on the bottom (Full connection). The terms in parentheses after 'Inception' are identifiers of the inception modules used in the original paper (Szegedy, Liu, Jia, Sermanet, Reed, Anguelov, Erhan, Vanhoucke, and Rabinovich, 2015). In this research, we attempt to evaluate several types of configurations by removing some layers from the original GoogLeNet. Here, • indicates that the layer was not included in the CNN.

Layer type	(The total number of layers) The number of outputs/Patch size/Stride				
	(27)	(23)	(17)	(11)	(6)
Convolution	16/7/3	16/7/3	16/7/3	16/7/3	16/7/3
Max pooling	16/3/2	16/3/2	16/3/2	16/3/2	16/3/2
Convolution	16/1/1	16/1/1	16/1/1	16/1/1	16/1/1
Convolution	48/3/1	48/3/1	48/3/1	48/3/1	48/3/1
Max pooling	48/3/2	48/3/2	48/3/2	48/3/2	•
Inception (3a)	64/*/*	64/*/*	•	64/*/*	•
Inception (3b)	120/*/*	120/*/*	•	120/*/*	•
Max pooling	120/3/2	120/3/2	•	•	•
Inception (4a)	128/*/*	128/*/*	128/*/*	•	•
Inception (4b)	128/*/*	•	128/*/*	•	•
Inception (4c)	128/*/*	•	128/*/*	•	•
Inception (4d)	132/*/*	132/*/*	132/*/*	•	•
Inception (4e)	208/*/*	208/*/*	208/*/*	•	•
Max pooling	208/3/2	208/3/2	•	•	•
Inception (5a)	208/*/*	208/*/*	•	•	•
Inception (5b)	256/*/*	256/*/*	•	•	•
Average pooling	256/5/1	256/5/1	208/21/1	120/21/1	48/43/1
Full connection (40% dropout)	2/-/-	2/-/-	2/-/-	2/-/-	2/-/-

¶ The activation function in each convolution layer is a rectified linear function and the activation functions in the fully connected output layers are softmax functions.

$*$) In the inception module, the input goes to all of the four paths – 1) 1×1 convolution, 2) 1×1 convolution and 3×3 convolution, 3) 1×1 convolution and 5×5 convolution, 4) 3×3 max pooling and 1×1 convolution – and outputs from these paths are concatenated as a multi-channel image after the operation. The values of the stride are 1 for all inception modules. Note that the inception module is counted as two layers.

and the quasi-minimal energy after step (4c) was $E = 6.82 \times 10^5$. Figure 2 shows the obtained correspondence between financial ratios and pixels under the 'Correlated' method, in which the digits represent the numbers attached to financial items in table 2 and the upper and lower rows in each cell represent the numerator and the denominator of financial ratios, respectively. The reason why the number of generated financial ratios is significantly small relative to the number of financial items is mainly because ratios whose denominators are zero for a certain company are excluded in step (2) of section 3.1.

Figure 3 shows examples of the generated images which are enlarged to a size of 256×256 pixels by using the nearest neighbor method so that they are suitable as inputs for GoogLeNet. Each subregion of 10×10 pixels corresponds to a specific financial ratio. When a financial ratio has the same value as the average over the 2062 companies in the reference dataset, the brightness value of its corresponding region is 128. Regions whose brightness values are close to white indicate that the value of the corresponding financial ratio is greater than the average, and conversely darker regions indicate that the value of the corresponding financial ratio is less than the average. Figure 3(a) shows the case when the correspondence between the financial ratios and pixels is determined at random and Fig. 3(b) shows the case when financial ratios with similar absolute correlation coefficients are placed closer together. The two images in the left column are of the same company belonging to the bankrupt enterprises class. Similarly, the two images in the right column are of a company belonging to the continuing enterprises class. Comparing these images, we can see results of arranging correlated financial ratios closer together.

4.2. Identification performance

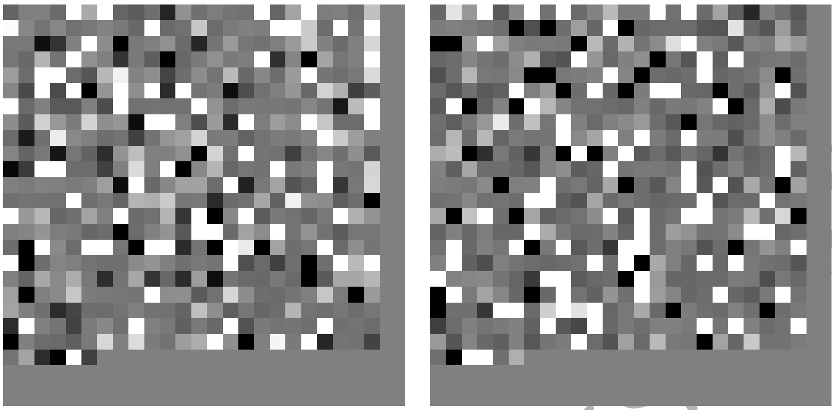
Table 3 shows the identification rates defined as correct estimation rates for each test dataset in the five patterns described in section 3.4. Table 3(a) shows the case when the correspondence between the financial ratios and pixels is determined at random and Table 3(b) shows the case when the correlated financial ratios are placed closer together. The identification rates are shown

Table 2: Financial items in our study under the parameter $p = 0.8$. Of the total of 263 financial items, those with a missing rate of $p = 0.8$ or more have been excluded, leaving 133 items in this list.

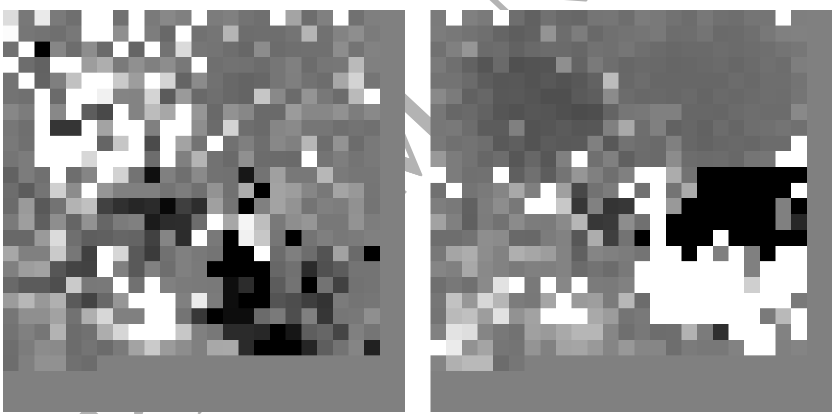
Balance Sheet (Assets Section)			
1	Current assets	24	Mechanical equipment and vehicles
2	Cash and cash equivalents	25	Tools, instruments and fixtures
3	Notes and accounts receivable	26	Other depreciable tangible fixed assets
4	Notes receivable	27	Construction in progress
5	Accounts receivable	28	Land and other non-depreciable tangible fixed assets
6	Securities	29	Intangible assets
7	Inventory	30	Software
8	Products	31	Goodwill
9	Real estate for sale	32	Other intangible assets
10	Semi-finished products or work in process	33	Investment and other fixed assets
11	Work in process	34	Investment securities, Affiliate stock, Affiliate investments
12	Raw materials	35	Investment securities
13	Prepaid expenses	36	Affiliate stock
14	Other accounts receivable	37	Affiliate investments
15	Short-term loans receivable	38	Long-term loan receivable
16	Short-term loans not for employees	39	Bankruptcy claim, rehabilitation claim
17	Deferred tax asset in current assets	40	Long-term prepaid expenses
18	Other current assets	41	Assets related to retirement benefits
19	Allowance for doubtful accounts in current assets	42	Guarantee deposit
20	Fixed assets	43	Deferred tax assets in fixed assets
21	Tangible fixed assets	44	Other fixed assets
22	Depreciable tangible fixed assets	45	Allowance for doubtful accounts in fixed assets
23	Buildings and structures	46	Total assets
Balance Sheet (Liability Section)			
47	Current liabilities	66	Other short-term allowance
48	Notes and accounts payable	67	Other current liabilities
49	Notes payable	68	Fixed liability
50	Accounts payable	69	Long-term debt, corporate bonds, convertible bonds
51	Short-term loan and bond payable	70	Corporate bonds, convertible bonds
52	Total debt repayment within one year	71	Corporate bonds
53	Short-term borrowings	72	Convertible bonds
54	Commercial Paper	73	Long-term debt
55	Long-term debt repayment within one year	74	Long-term accrued accounts
56	Bond debt, convertible bonds within one year	75	Total allowance
57	Bond debt within one year	76	Liabilities related to retirement benefits
58	Accrued accounts and expenses	77	Reserve for retirement benefits for officers
59	Accrued accounts	78	Other long-term allowance
60	Accrued expenses	79	Deferred tax liabilities in fixed liabilities
61	Accrued income tax	80	Deferred tax liability on revaluation
62	Accrued consumption tax	81	Asset retirement obligation
63	Advance payments received	82	Other fixed liabilities
64	Deposit received	83	Total liabilities
65	Accrued bonuses		
Balance Sheet (Equity Section)			
84	Shareholders' equity	90	Valuation difference on available-for-sale securities
85	Capital stock	91	Land valuation difference
86	Capital surplus	92	Foreign currency translation adjustment account
87	Retained earnings	93	Non-controlling shareholders' equity
88	Treasury stock	94	Capital attributable to parent company shareholders
89	Valuation, translation adjustments and others	95	Net assets
Profit-and-Loss Statement			
96	Sales volume	115	Assets disposal gains
97	Cost of sales	116	Gain on disposal of tangible fixed assets
98	Operating expenses	117	Gain on disposal of real estate
99	Gross profit	118	Other extraordinary income
100	Selling and general administrative expenses	119	Extraordinary expenses
101	Operating profit	120	Reorganization related loss
102	Non-operating income	121	Impairment loss
103	Interest income and dividends	122	Loss on valuation of securities
104	Interest income	123	Assets disposal loss
105	Dividend income	124	Loss on disposal of tangible fixed assets
106	Exchange gain	125	Loss on disposal of real estate
107	Equity method investment profit	126	Other extraordinary loss
108	Other non-operating income	127	Net profit before taxes
109	Non-operating expenses	128	Income taxes etc.
110	Interest expense	129	Income tax, inhabitant tax, total business tax
111	Other non-operating expenses	130	Income tax adjustments
112	Ordinary profit	131	Net profit
113	Extraordinary income	132	Net profit attributable to non-controlling shareholders
114	Gains on sales of securities	133	Net profit attributable to parent company shareholders

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
0	64 125	31 102	130 125	72 125	62 125	114 125	123 125	37 125	126 125	60 125	115 125	116 125	13 125	56 109	16 117	60 117	25 117	78 117	75 117	76 117	16 125	15 125	31 129	93 128
1	64 117	41 125	107 125	24 125	124 125	111 125	119 125	92 125	113 125	91 125	121 125	38 125	30 125	56 117	15 117	79 117	93 117	29 117	28 117	44 117	108 117	42 117	48 50	27 128
2	54 117	78 125	45 125	76 125	79 125	65 125	12 125	43 125	27 125	10 125	80 125	118 125	57 117	117 125	90 117	127 117	33 117	128 117	23 117	99 117	100 117	7 117	8 117	9 102
3	39 125	36 125	77 125	109 125	75 125	23 125	86 125	128 125	29 125	11 125	89 125	106 125	42 125	58 117	35 117	101 117	112 117	20 117	98 117	65 117	96 117	103 117	51 117	89 117
4	120 125	63 125	132 125	22 125	21 125	17 125	85 125	105 125	108 125	7 125	32 125	31 125	66 125	82 117	61 117	131 117	129 117	17 117	46 117	102 117	21 117	18 117	53 117	66 117
5	19 125	71 125	44 125	93 125	110 125	28 125	101 125	94 125	95 125	102 125	8 125	88 125	81 125	34 117	87 117	133 117	43 117	1 117	22 117	83 117	97 117	67 117	52 117	38 117
6	82 125	25 125	68 125	90 125	20 125	129 125	127 125	96 125	84 125	99 125	100 125	104 125	14 117	31 117	4 117	5 117	84 117	95 117	47 117	68 117	69 117	73 117	55 117	92 117
7	74 125	70 125	55 125	133 125	131 125	5 125	112 125	3 125	97 125	98 125	103 125	4 125	117 116	30 117	49 117	2 117	94 117	3 117	119 117	85 117	111 117	27 117	70 117	26 117
8	26 125	56 125	52 125	69 125	46 125	1 125	61 125	34 125	48 125	87 125	49 125	58 125	6 117	63 117	105 117	50 117	32 117	48 117	11 117	110 117	86 117	109 117	121 117	70 127
9	24 102	54 125	51 125	83 125	73 125	2 125	47 125	33 125	50 125	70 109	122 125	96 97	59 125	59 117	80 117	88 117	36 117	24 117	12 117	10 117	71 117	113 117	70 101	71 127
10	27 102	54 130	57 125	53 125	71 109	46 20	18 125	35 125	67 125	96 47	98 97	57 109	3 5	70 133	70 131	117 115	91 117	69 112	70 112	71 112	7 112	99 112	36 112	8 112
11	79 102	8 102	74 117	51 2	69 2	52 85	73 2	98 1	98 47	96 98	6 125	6 129	71 131	9 109	71 133	9 117	69 127	73 112	33 112	20 112	100 112	48 112	3 112	8 101
12	87 2	1 2	100 2	28 2	83 2	83 20	97 1	96 1	97 47	95 47	96 83	1 47	46 83	69 101	71 101	73 127	28 112	68 112	97 112	22 112	98 112	34 112	5 112	87 112
13	96 2	5 2	99 2	46 2	21 2	68 2	96 20	54 108	98 83	97 83	94 47	84 47	68 33	14 125	73 101	52 112	83 112	21 112	96 112	46 112	94 112	95 112	50 112	35 112
14	95 2	94 2	3 2	7 2	97 20	22 2	23 2	34 85	83 47	21 47	83 1	46 47	9 125	67 112	51 131	51 112	47 112	23 112	7 131	24 112	85 112	84 112	86 112	2 112
15	48 2	98 2	97 2	47 2	20 2	87 85	51 85	33 85	20 1	20 47	69 33	46 97	28 102	40 125	53 133	51 133	53 112	23 131	6 131	76 131	50 131	1 112	85 131	101 131
16	36 129	5 85	3 85	23 85	84 85	95 85	49 108	98 20	46 1	46 96	22 100	46 98	6 128	53 131	52 131	52 133	47 131	21 131	22 131	1 131	5 131	2 131	94 131	86 131
17	84 2	100 85	99 85	22 85	47 85	20 85	54 103	97 95	21 99	96 95	69 129	82 129	71 129	67 131	69 133	55 133	83 131	28 131	12 131	75 131	35 131	3 131	34 131	84 131
18	33 2	7 85	1 85	83 85	21 85	94 85	22 99	20 99	97 94	1 94	1 95	73 129	20 95	95 84	69 131	18 131	68 131	20 131	97 131	46 131	48 131	87 131	95 131	58 133
19	34 2	68 85	46 85	48 85	28 85	46 99	83 99	98 95	98 94	96 94	46 94	27 129	21 87	20 87	73 131	73 133	8 131	24 131	96 131	98 131	33 131	58 131	24 101	48 101
20	2 85	98 85	96 85	1 99	84 99	95 94	68 99	47 99	83 94	83 95	46 95	1 84	94 84	91 108	95 87	94 87	55 131	47 87	112 131	99 131	100 131	127 131	98 87	36 101
21	36 102	97 85	73 85	69 85	95 99	80 108	87 99	70 129	97 84	20 94	96 84	20 84	98 84	46 84	83 84	84 87	83 87	1 87	46 87	96 87	29 131	36 131	50 101	97 87
22	86 85	98 99	97 99	96 99	93 129	94 99																		
23																								

Figure 2: Correspondence between financial ratios and pixels under the 'Correlated' method. The digits represent the numbers attached to financial items in table 2, and the upper and lower rows in each cell represent the numerator and the denominator of financial ratios, respectively. The ones shown in red correspond to representative profitability indicators such as $(x, y) = (19, 18)$:ROA, (23,17):ROE, (18,19):Sales-profit ratio, (9,16):Total asset turnover, (3,22):Gross profit ratio, and representative safety indicators such as (11,12):Current ratio, (11,21):Fixed ratio, (10,20):Capital-to-asset ratio, although the numerator and denominator are reversed from the original definition in some cases.



(a) Correspondence between the financial ratios and pixels under the 'Random' method



(b) Correspondence between financial ratios and pixels under the 'Correlated' method

Figure 3: Examples of expressing financial ratios as an image: The left and right columns are for a single bankrupt enterprise and a single continuing enterprise, respectively. The image size has been enlarged to 256×256 pixels and each financial ratio corresponds to a subregion of 10×10 pixels. The size of the input images in GoogLeNet is fixed at 256×256 pixels, and on the bottom and right-hand edges of the images there are regions with a brightness value of 128 not corresponding to any specific financial ratios.

separately depending on the true classes.

It can be seen that the identification results are in most cases biased toward the continuing enterprises class. This is thought to be because the continuing companies are diverse in their financial conditions and so occupy a large portion of the feature space. It is also thought that some of the continuing companies are close to bankruptcy, which might lead to the identification boundaries moving to reduce the volume of the bankrupt enterprises class in the feature space.

To assess the identification performance with a single criterion, F-measures are derived by taking the bankrupt enterprises as positive instances and continuing enterprises as negative instances. This reflects the fact that the aim of this research can be regarded as detecting bankrupt companies from other enterprises. The results are depicted in Fig. 4. The reason for the low F-measures is that most of the data in the test set belong to the continuing enterprises class and thus the precision tends to be low. The result suggests that allocating neighboring pixel positions to highly correlated financial ratios improves the identification performance. However, even a random correspondence can take into account the relationships between distant pixels by making the CNN multi-layered. As a result, it seems that the performance difference between the two methods could be reduced to around 10% when the number of layers is 11 or more. Also, as in many problems such as image recognition, we can see that the performance improves as the number of layers increases.

We investigate the identification performance for each reason of delisting that can be deemed bankruptcy. Breakdown of the delisting reasons for our 102 bankrupt companies is 1) 72.55% of bankruptcy/rehabilitation/reorganization procedures, 2) 1.96% of excessive debt, 3) 1.96% of suspension of bank transactions, and 4) 23.53% of termination of business activities. On the other hand, the proportion of each delisting reason for 62 misidentified cases over the five test datasets with the 23 layers and 'Correlated' method is 1) 62.9% of bankruptcy/rehabilitation/reorganization procedures, 2) 1.61% of excessive debt, 3) 3.23% of suspension of bank transactions, and 4) 32.26% of termination of business activities. The second and third reasons are excluded from

Table 3: Identification rates (correct estimation rates) for each class with respect to the number of layers in the network configurations. The overall trend is that 1) the identification rate for the continuing enterprises class is higher, and 2) the identification rate improves as the number of layers is increased.

(a) Correspondence between financial ratios and pixels for the 'Random' method

Dataset not used for learning	True class	Total number of layers				
		6	11	17	23	27
Dataset A	Bankrupt	0.830	0.852	0.852	0.830	0.773
	Continuing	0.807	0.885	0.914	0.925	0.920
Dataset B	Bankrupt	0.784	0.920	0.898	0.920	0.909
	Continuing	0.862	0.837	0.891	0.899	0.883
Dataset C	Bankrupt	0.727	0.909	0.920	0.830	0.875
	Continuing	0.839	0.872	0.897	0.905	0.907
Dataset D	Bankrupt	0.682	0.852	0.909	0.830	0.875
	Continuing	0.834	0.919	0.875	0.907	0.898
Dataset E	Bankrupt	0.830	0.920	0.909	0.898	0.898
	Continuing	0.838	0.897	0.868	0.909	0.890
Average	Bankrupt	0.770	0.891	0.898	0.861	0.866
	Continuing	0.836	0.882	0.889	0.909	0.900

(b) Correspondence between financial ratios and pixels for the 'Correlated' method

Dataset not used for learning	True class	Total number of layers				
		6	11	17	23	27
Dataset A	Bankrupt	0.727	0.818	0.795	0.830	0.795
	Continuing	0.922	0.921	0.927	0.929	0.953
Dataset B	Bankrupt	0.830	0.864	0.886	0.932	0.909
	Continuing	0.882	0.879	0.900	0.884	0.903
Dataset C	Bankrupt	0.761	0.727	0.727	0.761	0.864
	Continuing	0.899	0.917	0.906	0.925	0.930
Dataset D	Bankrupt	0.716	0.886	0.898	0.875	0.875
	Continuing	0.897	0.923	0.930	0.949	0.925
Dataset E	Bankrupt	0.818	0.864	0.886	0.898	0.909
	Continuing	0.916	0.868	0.894	0.895	0.868
Average	Bankrupt	0.770	0.832	0.839	0.859	0.870
	Continuing	0.903	0.902	0.911	0.916	0.916

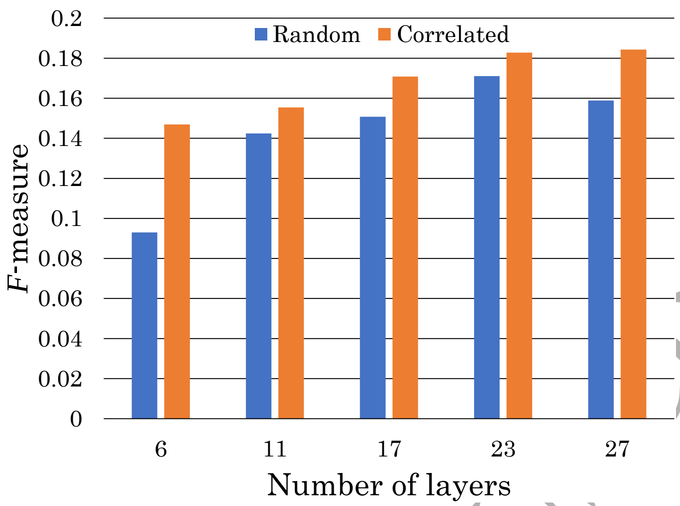


Figure 4: F-measures with respect to the number of layers. For all cases, the correspondences between financial ratios and pixel positions generated by the 'Correlated' method give us more favorable results than those by the 'Random' method.

consideration because their corresponding data are relatively uncommon. For the remaining two reasons, statistical hypothesis testing has been conducted by setting the null hypothesis that the ratio of the delisting reasons to the total number of misclassified test data is the same as the population ratio. The obtained p-values are 0.0444 for bankruptcy/rehabilitation/reorganization procedures and 0.0526 for termination of business activities, which implies that our obtained model can be slightly affected by considering the different four delisting reasons as bankruptcy, and there is a possibility of improving identification performance by eliminating companies that were delisted due to termination of business activities from datasets.

4.3. Influence of the number of training data points

As mentioned above, there are some differences between the identification rates for bankrupt and continuing enterprises. In order to reduce this imbalance, we attempt to reduce the size of the training dataset for the continuing enterprises class. From this section onward, only the 'Correlated' method for determining the correspondence between financial ratios and pixels is considered, along with a network consisting of 23 layers.

The number of continuing companies included in groups A-E is reduced to 15 or 10. Thus, the size of the continuing enterprises class in each training dataset becomes 15 companies $\times$ 4 sets $\times$ (4 periods + 90 synthetic values) or 10 companies $\times$ 4 sets $\times$ (4 periods + 90 synthetic values). Companies removed from the training dataset are added to the test dataset. The results with these new datasets are shown in Table 4. As the number of continuing companies is reduced, the identification performance of bankrupt enterprises rises, as expected, and that for continuing enterprises decreases. The differences between the average identification rates are smallest when 15 continuing companies are included in groups A-E.

The F-measures are obtained in the same way as before, and are 0.138 for the 15-company case and 0.096 for the 10-company case. This indicates that the highest performance with the maximum F-measure is achieved when the numbers of data points in the two classes are equal. However, from a practical point of view, it may be preferable to ensure the identification rate for the bankrupt enterprises class (i.e., recall when detecting companies that will become bankrupt) is above a certain level. There is no general method for determining the optimal size of the training dataset and it is thought that analysts should adaptively determine the size according to the problem setting and the required performance level.

4.4. Influence of synthetic financial data generated by the weighted average

It is possible to use other approaches to reduce the size of the training dataset for the continuing enterprises class. The reason for limiting the number of continuing companies in each of groups A-E to 20 was to make the number equal to the number of bankrupt companies. However, much data on continuing companies are available and thus the real data can be used without having to rely on the weighted average operation to increase the number of training data points.

Therefore, 150 continuing companies are randomly selected for each of groups A-E, and only four periods' real data from these companies are incorporated in the datasets A-E, whereby the data of bankrupt companies are unaltered (including both real and synthetic data). Note that the training data belonging to the continuing enterprises class do not include the synthetic data. The test datasets are also changed so that the continuing companies to be learned are not included.

Table 5 shows the results of experiments based on these modified datasets. The identification rates for continuing companies are approximately the same as in Table 3 but those for bankrupt companies are significantly improved. The F-measure is also improved at 0.189. Compared to the 7520 training data points for the bankrupt enterprises class, only 600 training data points are utilized for the continuing enterprises class, which implies that the simple weighted average operation does not generate inherently different data and it does not have the same effect as increasing the number of real data points.

4.5. Comparison of the identification performance with other methods

The proposed method is compared with other representative machine learning methods. Here, comparisons are made with the algorithms for 1) classification and regression trees (CART), 2) linear discriminant analysis (LDA), 3) support vector machine (SVM), 4) multi-layer perceptrons (MLP), and 5) AdaBoost, using our financial ratios as feature vectors for our five training and test datasets. However, due to computational impossibility for missing values, financial ratios with missing values are excluded in advance. An overview of the use of these algorithms for bankruptcy prediction is given below.

CART This is a representative algorithm for decision trees. The financial ratio that minimizes the Gini coefficient is selected for determining each new partition. The maximum depth of a tree is set to 30. However, once there is only data belonging to a single class in a node, no further divisions can be made from there.

Table 4: Identification rates when the continuing enterprises included in groups A-E are reduced to 15 and 10 companies. These results are obtained under the 'Correlated' method for determining the correspondence between the financial ratios and pixels, and using 23 layers for the network configuration. The number of training data points for bankrupt enterprises class remains at 7520 but that of continuing enterprises class decreases to 5640 and 3760 for 10 and 15 continuing companies in each group, respectively.

Dataset not used for learning	True class	Number of training data points for the continuing class	
		5640	3760
Dataset A	Bankrupt	0.784	0.886
	Continuing	0.915	0.837
Dataset B	Bankrupt	0.920	0.955
	Continuing	0.853	0.789
Dataset C	Bankrupt	0.875	0.898
	Continuing	0.870	0.817
Dataset D	Bankrupt	0.852	0.943
	Continuing	0.909	0.801
Dataset E	Bankrupt	0.920	0.909
	Continuing	0.863	0.822
Average	Bankrupt	0.870	0.918
	Continuing	0.882	0.813

Table 5: Identification rates when each training dataset contains only real data in regard to the continuing enterprises class. These results are obtained under the 'Correlated' method for determining the correspondence between financial ratios and pixels and using the 23 layers for the network configuration. The number of training data points for the bankrupt enterprises class remains at 7520 (including both real and synthetic data) but that of the continuing enterprises class is only 600 (only real data).

Dataset not used for learning	True class	Number of training data points for the continuing class
		600
Dataset A	Bankrupt	0.841
	Continuing	0.932
Dataset B	Bankrupt	0.943
	Continuing	0.879
Dataset C	Bankrupt	0.920
	Continuing	0.918
Dataset D	Bankrupt	0.841
	Continuing	0.929
Dataset E	Bankrupt	0.898
	Continuing	0.910
Average	Bankrupt	0.889
	Continuing	0.914

LDA Financial ratios that contribute more to the decrease in the Gini coefficient for the aforementioned CART algorithm are sequentially added to the feature vectors. Discriminant functions are repeatedly derived as financial ratios are sequentially added. A method of carrying out discriminant analysis using variables selected based on a decision tree was proposed by Shirata (1998) in the context of bankruptcy prediction. However, Shirata also used her knowledge of accounting and finance as well as knowledge gained from preceding research.

SVM As with LDA, the financial ratios with the greatest contribution to the decrease in the Gini coefficient are sequentially applied in the discrimination with the Gaussian kernel. Parameters for adjusting the balance between the soft margin and the margin maximization and parameters for the variance in the Gaussian kernel are optimized via a grid search so that the maximum average identification rate for the test datasets can be obtained.

MLP Unlike CNNs, MLPs consist of only fully connected layers. The optimal parameters for the number of hidden layers (between 2 to 6), the mini-batch size, and the learning coefficients are searched for so that the maximum average identification rate for the test datasets can be obtained. Our method outperforming the MLP approach would show the validity of the utilization of the CNN.

AdaBoost Regarding each of the financial ratios as a weak classifier with a simple threshold, the best weak classifier and the optimal value of its threshold are obtained at each boosting step. This is equivalent to setting a decision tree of depth 1, i.e., a decision stump, as the base-classifier of AdaBoost. The weights of the training data misidentified by the extracted weak classifier are increased at the next step. The weighted majority voting of the sequentially selected weak classifiers becomes the final (strong) classifier to carry out the identification. This method was proposed by Takata, Hosaka, and Ohnuma (2015) and was then extended to the utilization of the RealAdaBoost (Hosaka and Takata, 2016) and to early bankruptcy predictions (Takata, Hosaka, and Ohnuma, 2017).

Average identification rates obtained via the methods described above and with our proposed method (in the case of the 'Correlated' method and using 23 layers) are shown in Fig. 5. All methods use the same dataset as described earlier. The colors of the lines represent the different methods, and Figs. 5(a) and 5(b) exhibit the average identification rate over the five test datasets for bankrupt enterprises and for continuing enterprises, respectively. The horizontal axis represents the number of financial ratios incorporated into the discriminant function through the sequential feature selection. However, CART, MLP and our proposed method do not involve feature selections and so a constant value is presented regardless of the value on the horizontal axis. Although the horizontal axis is shown up to a value of 50, there are no large fluctuations in the identification rate beyond this.

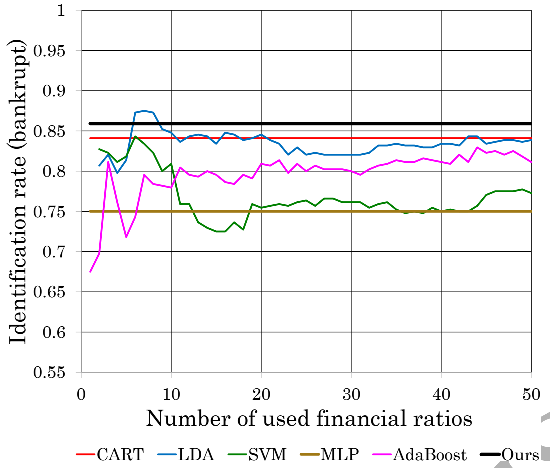
The proposed method shows a notably higher identification performance for continuing enterprises compared to the other methods. The identification rates of the LDA method are higher than that of our proposed method for bankrupt enterprises in some regions but this is because the identification by LDA is greatly biased toward the bankrupt enterprises class. To compare two kinds of identification rates simultaneously, an ROC (Receiver Operating Characteristic) curve is obtained by varying the threshold for the predictive probability or the discriminant function finally derived with these methods. The ROC curves, when the number of used financial ratios is set to 50, are depicted in Fig. 6, including the curve of Altman's Z''-score (Altman and Hotchkiss, 2006; Altman, Danovi, and Falini, 2013) which will be described immediately below. The black circle on the curve of the proposed method corresponds to the result shown in table 3(b), which corresponds to the simplest case of classifying a target company to the class with the higher predictive probability. Considering that the curve passing through the upper left (a region of high true positive rate and low false positive rate) is desirable, the superiority of the proposed method to conventional ones is obvious although there are some relatively inferior areas in the upper right region.

Regarding the performance of AdaBoost, it can be said that the identification rates for both classes are reasonably high and well balanced particularly after about 10 iterations. This implies that favorable performance can be obtained simply by setting the number of iterations at more than a certain value. Considering the fact that this AdaBoost operation does not have other parameters to tune, thanks to using a highly simple decision stump as base weak classifiers, this method is extremely easy to implement in practical situations.

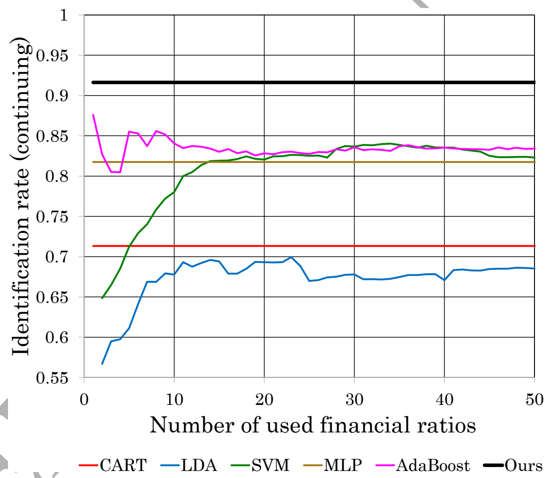
It is also significant to compare our results with the traditional financial indicator of bankruptcy prediction. Here, we consider Altman's Z''-score (Altman and Hotchkiss, 2006; Altman, Danovi, and Falini, 2013) which is described as

$$\begin{aligned} Z'' &= 6.56X_1 + 3.26X_2 + 6.72X_3 + 1.05X_4, \qquad (5) \\ & \quad X_1 : \text{Working Capital / Total Assets}, \\ & \quad X_2 : \text{Retained Earnings / Total Assets}, \\ & \quad X_3 : \text{Earnings before Interest and Tax / Total Assets}, \\ & \quad X_4 : \text{Book Value Equity / Total Liabilities}. \end{aligned}$$

When the obtained score is less than the threshold which should be determined according to the target sectors or countries, the firm is recognized as the default situation. For our 102 bankrupt enterprises and 2062 continuing enterprises (four years data per company), we set various values of the threshold to generate the ROC curve, which is also depicted in Fig. 6 with other conventional methods. We can see that the Z''-score is inferior not only to our proposed method but also to other methods, which implies that the four indicators used in Z''-score are not necessarily effective for Japanese bankruptcy prediction. This inference is also supported by the fact that LDA in our experiments using financial ratios adaptively chosen by the CART-based variable selection actually improves the performance.



(a) Identification rate for the bankrupt enterprises class with respect to the number of financial ratios incorporated in the discriminant function.



(b) Identification rate for the continuing enterprises class with respect to the number of financial ratios incorporated in the discriminant function.

Figure 5: Comparison experiments. Average identification rates ((a): for bankrupt enterprises, (b): for continuing enterprises) by CART, LDA, SVM, MLP, and AdaBoost as well as by our proposed method are illustrated. The horizontal axis represents the number of financial ratios incorporated into the discriminant function and the vertical axis represents the average identification rate over the five test datasets. It can be observed that our proposed method outperforms the other machine learning algorithms.

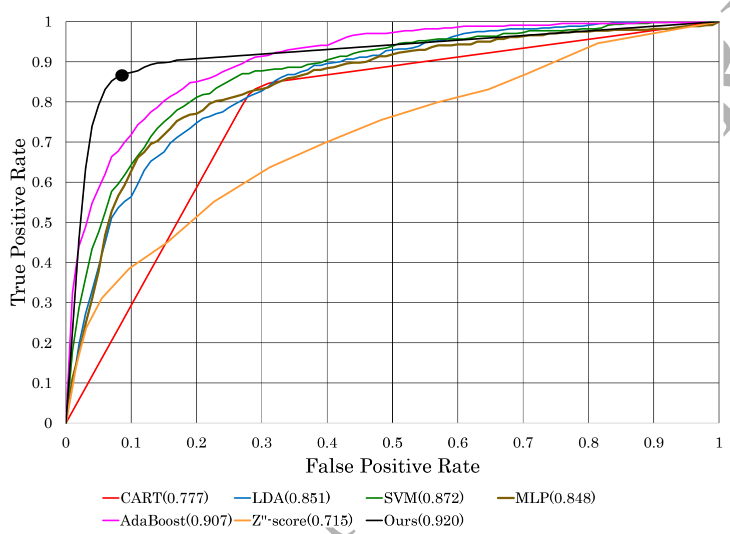


Figure 6: ROC curve for comparison with traditional methods. Regarding the bankruptcy as a positive instance, true positive rates are illustrated with respect to the false positive rates. These evaluations can be conducted by changing the threshold for predicted bankrupt probabilities (LDA, MLP, and ours) or for discriminant functions (SVM, AdaBoost, and the Z''-score). Note that since it is generally difficult to generate an ROC curve for CART, two trivial points located in both corners are connected with the point actually realized in the experiment. The numbers in the brackets in the legend represent the values of AUC (Area Under Curve). It can be easily confirmed that our proposed method outperforms the other conventional methods.

5. Conclusions

We have proposed a method for applying a CNN to bankruptcy prediction. In our method, a set of financial ratios are represented as a grayscale image where each financial ratio corresponds to a fixed pixel position, and the generated images are used as training data for a CNN based on GoogLeNet. A numerical evaluation revealed that allocating neighboring pixel positions to highly correlated financial ratios is more appropriate for our purpose than placing them at random. Our analysis also indicated that the proposed method outperforms representative conventional methods using CART, LDA, SVM, MLP, AdaBoost and Altman's Z''-score. Furthermore, the proposed method of conversion from financial ratios to an image has the potential to be applied to general numerical data in a variety of contexts other than bankruptcy prediction.

However, unlike some conventional methods, it is hard to know from the proposed method which of the financial ratios has a stronger impact on bankruptcy prediction. Therefore, we have to admit that the proposed method is not suitable for the purpose of investigating the causes of bankruptcy.

There are some remaining unresolved issues. For example, as mentioned in section 4.3, the impact of using different proportions of real and synthetic data for each class should be investigated in more detail. The network structure used in this research was based on GoogLeNet but this was not selected through in-depth theoretical consideration. It is necessary to verify whether other network configurations are more effective for the problem of bankruptcy prediction. Regarding the imaging of financial ratios, there is also a need to verify whether the identification performance could be improved by criteria other than the correlation coefficient used in this study.

Acknowledgement

This research was supported by JSPS KAKENHI grant number 15K21395.

References

Aggarwal, S., Aggarwal, S., 2016. Deep investment in financial markets using deep learning models. International Journal of Computer Applications 162, 403–413.

Alfalo, E., García, N., Gámez, M., Elizondo, D., 2008. Bankruptcy forecasting: An empirical comparison of adaboost and neural networks. Decision Support Systems 45 (1), 110–122.

Altman, E. I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. The Journal of Finance 23 (4), 589–609.

Altman, E. I., Danovi, A., Falini, A., 2013. Z-score models' application to italian companies subject to extraordinary administration. Journal of Applied Finance 23 (1), 1–10.

Altman, E. I., Hotchkiss, E., 2006. Corporate Financial Distress and Bankruptcy, 3rd Edition. Wiley.

Back, B., Laitinen, T., Sere, K., 1996. Neural networks and genetic algorithms for bankruptcy predictions. Expert Systems with Applications 11 (4), 407–413.

Bao, W., Yue, J., Rao, Y., 2017. A deep learning framework for financial time series using stacked autoencoders and long-short term memory. PLoS One 12 (7).

Brédart, X., 2014. Bankruptcy prediction model using neural networks. Accounting and Finance Research 3 (2), 124–128.

Bryant, S. M., 1998. A case-based reasoning approach to bankruptcy prediction modeling. Intelligent Systems in Accounting, Finance and Management 6 (3), 195–214.

Chong, E., Han, C., Park, F. C., 2017. Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. Expert Systems with Applications 83, 187–205.

Dahl, G. E., Yu, D., Deng, L., Acero, A., 2012. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. IEEE Transactions on Audio, Speech, and Language Processing 20 (1), 30–42.

Ding, X., Zhang, Y., Liu, T., Duan, J., 2015. Deep learning for event-driven stock prediction. Proceedings of the 24th International Joint Conference on Artificial Intelligence, 2327–2333.

Gordini, N., 2014. A genetic algorithm approach for smes bankruptcy prediction: Empirical evidence from italy. Expert Systems with Applications 41 (14), 6433–6445.

Graves, A., Mohamed, A., Hinton, G., 2013. Speech recognition with deep recurrent neural networks. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 6645–6649.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. IEEE Conference on Computer Vision and Pattern Recognition, 770–778.

Hosaka, T., Takata, Y., 2016. Corporate bankruptcy forecast using realadaboost. INFORMATION -An International Interdisciplinary Journal 19 (6(B)), 2285–2298.

Irsoy, O., Cardie, C., 2014. Deep recursive neural networks for compositionality in language. Proceedings of the International Conference on Neural Information Processing Systems, 2096–2104.

Kiros, R., Salakhutdinov, R., Zemel, R., 2014. Multimodal neural language models. Proceedings of the 31st International Conference on Machine Learning 32 (2), 595–603.

Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012. Imagenet classification with deep convolutional neural networks. Proceedings of Neural Information Processing Systems.

Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE 86 (11), 2278–2324.

Lee, J., Jang, D., Park, S., 2017. Deep learning-based corporate performance prediction model considering technical capability. Sustainability 9 (6), 899(article number).

Li, H., Sun, J., 2009. Predicting business failure using multiple case-based reasoning combined with support vector machine. Expert Systems with Applications 36 (6), 10085–10096.

Lin, M., Chen, Q., Yan, S., 2013. Network in network. arXiv:1312.4400.

Mohamed, A., Dahl, G. E., Hinton, G., 2012. Acoustic modeling using deep belief networks. IEEE Transactions on Audio, Speech, and Language Processing 20 (1), 14–22.

Odom, M. D., Sharda, R., 1990. A neural network model for bankruptcy prediction. Proceedings of International Joint Conference on Neural Networks, 163–168.

Ohlson, J. A., 1980. Financial ratios and the probabilistic prediction of bankruptcy. Journal of Accounting Research 18 (1), 109–131.

Persio, L. D., Honchar, O., 2017. Artificial neural networks architectures for stock price prediction: comparisons and applications. International Journal of Circuits, Systems and Signal Processing 10 (2), 40–43.

Ramakrishnan, S., Mirzaei, M., Bekri, M., 2015. Corporate default prediction with adaboost and bagging classifiers. Journal Teknologi (Sciences & Engineering) 73 (2), 45–50.

Sartori, F., Mazzucchelli, A., Gregorio, A. D., 2016. Bankruptcy forecasting using case-based reasoning: The creperie approach. Expert Systems with Applications 64 (C), 400–411.

Shin, K., Lee, T. S., Kim, H., 2005. An application of support vector machines in bankruptcy prediction model. Expert Systems with Applications 28 (1), 127–135.

Shirata, C. Y., 1998. Financial ratios as predictors of bankruptcy in japan: an empirical research. Proceedings of the Second Asian Pacific Interdisciplinary Research in Accounting Conference, 437–445.

Shirata, C. Y., Sakagami, M., 2008. An analysis of the going concern assumption: Text mining from japanese financial reports. Journal of Emerging Technologies in Accounting 5 (1), 1–16.

Siripurapu, A., 2015. Convolutional networks for stock trading. Stanford University Department of Computer Science, Course Project Reports.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.

Takata, Y., Hosaka, T., Ohnuma, H., 2015. Financial ratios extraction using adaboost for delisting prediction. Proceedings of the Seventh International Conference on Information, 158–161.

Takata, Y., Hosaka, T., Ohnuma, H., 2017. Boosting approach to early bankruptcy prediction from multiple-year financial statements. Asia Pacific Journal of Advanced Business and Social Studies 3 (2), 66–76.

Yao, K., Zweig, G., Hwang, M., Shi, Y., Yu, D., 2013. Recurrent neural networks for language understanding. Interspeech 2013.

Yeh, S., Wang, C., Tsai, M., 2015. Deep belief networks for predicting corporate defaults. Proceedings of the 24th Wireless and Optical Communication Conference, 159–163.

Yoshihara, A., Fujikawa, K., Seki, K., Uehara, K., 2014. Predicting stock market trends by recurrent deep neural networks. Proceedings of Pacific Rim International Conference on Artificial Intelligence (PRICAI), 759–769.