AI Based Advertisement Image Classification from Newspapers

Santhi. P

Assistant Professor, Department of Computer Science and Engineering (Data Science), IES College of Engineering, Kerala, India
Email_id: santhip@iesce.info

Abstract

News Papers has unavoidable role in the human society. The major outcomes of newspapers are day to day information, articles, advertisements etc. Among these, advertisements have significant influence in human life. From the very beginning of social life, man tried to advertise what he had, to others who were looking for. In the modern age advertising industry became so strong to influence our society and economy in several ways. Advertising helps to reach more people the messages that promote products and services. Reaching the target audience will improve the profit without wasting money or time. Newspapers are important factors that boost up advertising industry. Artificial Intelligence has significant role in solving problems related with computer vision and image processing which can be applied in real life situations. Various machine learning techniques and models used to make new predictions or findings from known/unknown sources of data. In this paper an effort is done to study and analyze the works related with advertisement classification from newspapers and magazines

Keywords: Newspapers, Advertisement classification, Advertising Industry, Artificial Intelligence, machine learning.

DOI: https://doi.org/10.5281/zenodo.15010959

1. Introduction

Newspapers are strong promoters of advertisements. Even though the digital marketing has gained a major position due to large use of internet, common people still depend Newspapers as it has become a part of their life when they wake up in the morning. Immense growth of internet not bound the newspapers as mere hardcopies, but e-newspapers are available which will make the people more comfortable to read the papers even in their office or in travel etc. Advertisements inspire the the people to buy products or services that enhance cash flow and which will make our economy dynamic. Advertisements mainly fall into the major classes like

1. Job
2. Education
3. Trade
4. Services
5. Politics &Administration
6. Finance

Producers and consumers looking forward where to find proper advertisements. Newspapers have some their own policies regarding the cost, space, size of the advertisements which will make the advertisers compelled to choose

some particular newspapers. So there are some visible differences among the newspapers while comparing the advertisements. Identifying the exact newspaper that matches the customer interest is very important. Similarly the advertisers also need to identify the place where there products or services should be displayed. In this context, understanding the trend among newspapers in choosing advertisements is relevant. A manual search of this trend is absolutely an impractical one. Here we enquire the possibility of AI applications.

1.1. Machine Learning

Humans learn from past experience. If humans can train machines to learn from past data and put use in a faster way, it is called machine learning. For this purpose, the machines need to be programmed using prior knowledge. We need to develop a model, and computer programs are used to optimize the model's parameters using training data. The model could be descriptive or predictive, make it enable to learn and do future predictions. The study of machine learning is what is referred to as "the field that makes computers to be able to learn without being explicitly programmed." There isn't a single definition of machine learning, though. The phrase has been defined variously by several authors

Learning Models

Machine learning focuses on leveraging the appropriate features to create models that accomplish the appropriate goals [13]. With or without prior information, a learning model can be educated from a set of inputs, and using that inference, the model can reason over previously unexplored data and generate predictions about that data. Three categories have been created from the fundamental concept of learning models. The sample space or instance space for a particular problem is the set of all potential outcomes. The different learning models are listed as follows.

- Putting an expression in logic. (Logical models)
- Making use of the instance space's geometry. (Geometric models)
- By classifying the instance space with probability. (Probabilistic model)

1.1.1. Supervised learning

Supervised learning makes use of a mapping function that identifies the correspondence between inputs and outputs using some sample input output pairs [13]. According to some training data with the right answers, the algorithm being generalized to respond appropriately to all possible inputs. It's also referred to as learning from examples. From tagged data, supervised learning gains knowledge. The data which has a particular value or label assigned to it is called labelled data. In supervised learning, a model is generated which is trained with a set of labelled inputs and the model will recognize certain pattern from each input and associate that pattern to label that we already provided. When we provide a new input, the model tries to recognize the pattern and its associated label then put that input into corresponding class. This type of learning process is known as classification. Another type of supervised learning in which the output variable is a numerical value rather than a categorical value. Such learning method is known as regression .Eg: Identifying the image of an apple is a classification method. Predicting the house price according to some patterns in the input is a regression method. Real life applications of supervised learning includes Spam detection, object classification, speech recognition, tumor identification etc.

1.1.2. Unsupervised learning

Unsupervised learning is used to reach into some conclusions from input data without labels. An observation's

classification or categorization is not present in unsupervised learning techniques. Inputs that share characteristics are grouped together by the algorithm, which seeks to find commonalities between them.[13]. Unsupervised learning works on unlabeled data without having any prior knowledge. The learning model identifies some interesting patterns from the inputs and these patterns are used to form clusters of output. Elements in each cluster are similar in attributes. But the clusters are different in their features. This type of learning is known as clustering. Eg. In search engines, whenever a user input a query then a number of websites details are shown which were attributed by that query. Unsupervised learning uses another methodology to identify the relation between data items. It is known as association mining. The concept of discovering association rules through market-basket data analysis, such as "A customer will have a probability p% of buying a product y, if he buys products x1 and x2. It establishes the group of items that co-occur in the collection. A practical illustration of unsupervised learning is the segmentation of the market based on consumer behavior.

1.2. Classification

Data classification is a crucial machine learning problem. According to the features extracted from training set of data that contains observations (or instances) whose category membership is already known; classification process determines a set of categories to which a new observation belongs. A classifier is an algorithm that performs classification. The classifier associates with a mathematical function which assigns input data to a category.

Important classification techniques

1.2.1 Naive bayes

Nave Bayes is a statistical learning method that, in accordance with the Bayes Theorem [10], forecasts the likelihood that a data sample will belong to a particular class where each class is formed according to some feature vector. The class with the highest probability is regarded as the most likely class. Each component that makes up the model is separate from the others. Because of the strong assumption of feature independence, Naive Bayes classifiers have less accuracy than other more refined learning algorithms. This is because it is not realistic for the features to be totally independent of one another. Comparatively speaking, it is a faster way of classification with less storage need. Naive Bayes classifiers can handle missing values quite well.

1.2.2 Support vector machine (svm)

It is a classification as well as regression prediction tool that improves forecast accuracy and avoids over fitting the data by applying machine learning concepts. A hypothesis space which consists of no of linear functions in high dimensional space is used and trained by using optimization theory based learning algorithms that incorporate a learning bias. A system that use such a hypothesis space and do the above task is called support vector machine [10]. SVMs are capable of performing a non-linear classification effectively by implicitly transforming their inputs into large feature spaces. Support vector machines have the following advantages [9].

- Effective in places with higher dimensions.
- It focuses on training optimization and makes use of multimodal classes.
- Uses a subset of training points in the decision and is very useful when the number of dimensions is considerably more than the number of samples.

1.2.3 k-nn (k-nearest neighbor)

The K-NN technique for picture classification is a supervised machine learning algorithm that can be implemented in a simple way [10]. The training phase is extremely straightforward and consists of storing the training image set's feature vectors and labels. During the initial phase of recognition, K-NN first fix the k-nearest data samples to the query data sample and then applies a class label that is most suitable for the k-nearest data samples, categorizing the unknown data samples based on its nearest neighbors whose categories are already known. When just a small fraction of characteristics are suitable for classification purposes, the KNN algorithm treats all features equally when computing similarity, which can result in misclassification (Kim1 et al., 2012).The advantage of these method is that minimum no of parameters(k) need to be tuned and there is no need of linearly separable classes.it also support multimodal classes. But some of the demerits of this method is it is difficult to set the value for K and it doesn't support higher dimensional feature space. Missing values are not replaced by this algorithm and storage requirements is slightly high.

1.2.4 Convolutional neural network (cnn) :

A deep learning architecture created for handling structured data is a convolutional neural network [12]. When analyzing input images with lines, gradients, circles, or even faces and eyes, CNN is particularly effective. Because of the special features of Convolutional neural networks, they are very useful in computer vision applications. CNN can operate directly on unprocessed photos. Using forward feed layers, a convolutional neural network feeds layers. Each of CNN's numerous hidden layers, which are stacked one on top of the other, is capable of recognizing more complex structures. It is feasible to distinguish between human faces and handwritten text using these convolutional layers. Convolutional layer, pooling layers, fully connected layers and flatten layers are the hidden layers. It is possible to identify different aspects in an image by using various filters. The characteristics generated through convolution are then subjected to a nonlinear activation function (ReLU). There is a pooling layer added after each convolutional layer. There are numerous ways to build upon this layer stack. A flattening layer is put after the final pooling layer. A fully connected layer is then set. The classification process is carried out, and classification output is acquired, through a series of fully connected layers. Three-dimensional neurons, such as those with width, height, and depth, are found in CNN layers. Sparse connectivity, shared weights, and pooling are its key characteristics. The two main benefits of CNN are that it is a speedier training process with less storage needed, and that it offers feature extraction capabilities in an automatic way without requiring advance knowledge of features. This classifier's drawback is that it requires computationally time-consuming training with sizable data sets. There are many modal parameters that must be tweaked, and network design problems are challenging to solve. It is also particularly sensitive to unimportant traits.

2. .Related works in literature

Even though there are so many works in the field of image processing, it could be found very few research works in the field of advertisement classification especially on newspapers, Unfortunately in Malayalam, such an attempt yet to be initiated.

A classification approach using Naïve Bayes classifier is used for advertisement classification [1].It used text contents of the advertisements for classification. The developed prototype consists of three modules. In the first module a classification model associates categories to advertisements. Second module is a tagging one that identifies

textual information that related with advertisement category. In the final phase a predefined form of each category filled with tagged information

There is an another attempt to classify the advertisements of newspapers and websites using rule based filters[2].Segmented images are applied with rule based filters and learning based filters to produce classification. Visual features and semantic features extracted from the images are used for classification.

Another method of categorizing internet advertisements based on CNN offers a binary classification, such as whether or not there is an advertising present [3]. The nLmF-CNN model is the one that was put forth in the work. The proper values of the parameters (n, m), are determined where the number of layers are denoted by n and number of filters are denoted by m. Images that have been captured online are used as input. Utilizing the ConvNetJs package, the processing parts of the nLmF-CNN are created as deep neural networks. The model offers a YES/NO binary categorization. YES indicates the presence of advertising. NO indicates that there are no advertising. To identify article types, text, lines, and drawings from articles are retrieved and applied with rule-based filters [4].

A method for newspaper layout analysis was discovered [5] by identifying linked elements like lines, text, and graphs and merging the related elements while taking qualities into consideration. In order to identify the geometric structure from an image, layout analysis is used. A novel homogeneity-related definition of distance is used using a component-based bottom-up algorithm, and it keep a dynamic minimum distance mechanism to determine the order in which the components merge. Another segmentation-based method for identifying patterns of components made up of neighboring rectangular sections was discovered [6]. The technique used in the work finds all lines in the document, even those that are joined to other parts. The image is segmented into patterns using a bottom-up method, and each pattern is then assigned to one of seven kinds.

To categorize the discovered blocks and ascertain the connection between neighboring blocks, a fixed point model was developed [7]. Based on where online images are located on a web page, a webpage segmentation method was presented [8] to extract web images and their contextual information.

A CNN-based model is suggested to recognize magazine advertisements or articles by learning visual elements from scanned images [9]. The three layers of the proposed architecture in the paper are convolutional, max-pooling, and fully connected layers (hidden and output layers). The job consists of two steps: feature extraction and categorization. To find the most accurate model, three different CNN models are applied to the same dataset of input images [10]. So many image processing classification techniques are found in literature. A brain image retrieval system using KNN classifier is implemented by extracting features like shape and texture with an accuracy of 52% [15]. Another attempt to classify images using Naïve Bayes method by extracting multiple features and find the optimal features is done by Kun Fang [16] . Roman numerals Recognition Application uses an ANN with a precision rate of 86% and a second powerful classifier support vector machine to compile all the classed results.[17] In literature [22], when classification is carried out on a common database using machine learning, a comparison of classifier algorithms' accuracy is made. Handwritten digital grayscale images of ten types are included, and examined 60,000 training photos having 28 by 28 resolution. Fig. 1 shows a summary of the results obtained from various classifier algorithms.

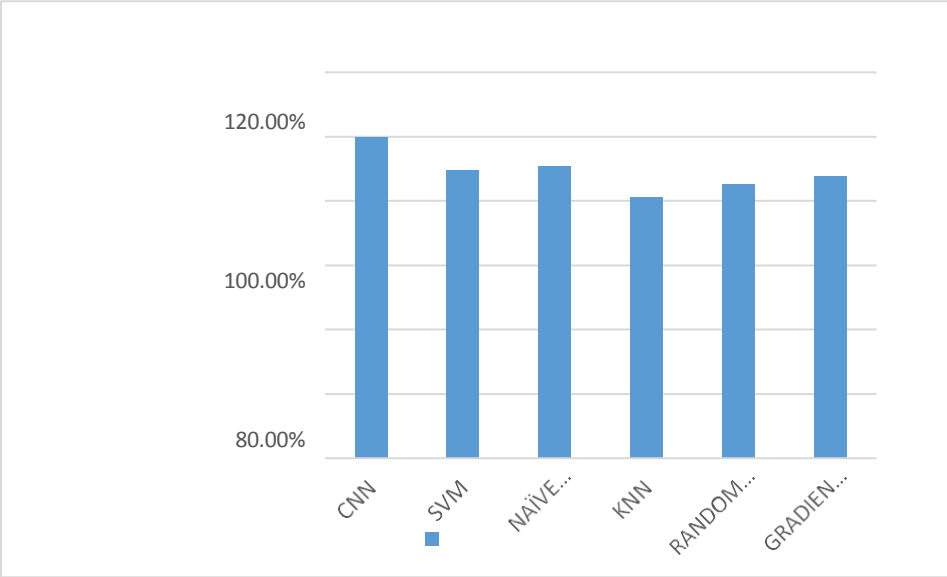

Figure 1: Comparison of different classifier algorithms

1. Proposed methodology

Since the breakthrough of CNNs with the ImageNet classification convolutional neural networks (CNNs) have become the gold standard for image recognition and classification tasks. This proposed research seeks to explore various CNN-based models to classify advertisement images collected from a variety of newspapers. Fig 2. Shows the architecture of a general CNN. The model design involves tuning several key hyper parameters to optimize the performance of the network for this specific task.

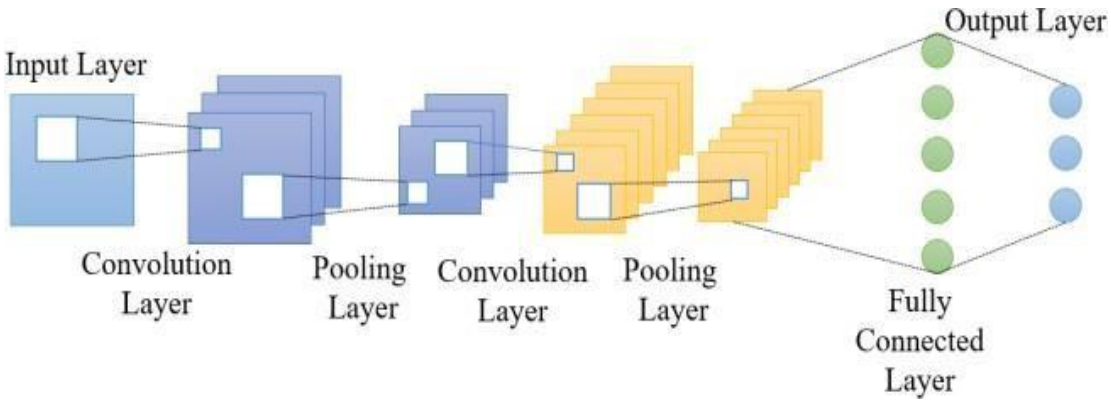

Figure 2: General CNN architecture with different layers

3.1. CNN Hyperparameters

In training a CNN for image classification, several hyperparameters must be carefully tuned to ensure optimal performance and accuracy. These include the batch size, number of epochs, optimization algorithm, learning rate, momentum, weight initialization techniques, activation functions, dropout regularization, and the number of neurons in each hidden layer. Below is a description of each hyperparameter:

1. Batch-size

- Definition: The batch size in CNN training refers to the number of training samples that are processed

before the model's internal weights are updated.

- Effect: CNNs are highly sensitive to the batch size, and selecting an appropriate value is crucial for training stability and model convergence. A smaller batch size may lead to a more noisy gradient, while a larger batch size can speed up training but may cause memory issues or overfitting.

2. Number of Epochs

- Definition: An epoch represents one complete cycle through the entire training dataset.
- Effect: The number of epochs defines how many times the model will see and learn from the training data. A sufficient number of epochs ensures that the model learns well, but too many epochs may cause overfitting.

3. Optimization Algorithm

- Definition: Optimization algorithms are used to minimize the loss function by adjusting the weights during training. Several algorithms are available to help CNNs converge quickly and efficiently.

4. Learning Rate and Momentum

- Learning Rate: This parameter controls the size of the step the optimizer takes when updating the model's weights after each batch.
  - Typical values: [0.001, 0.01, 0.1, 0.2, 0.3]. \
  - Effect: A high learning rate may lead to overshooting the optimal weights, while a low learning rate could slow down convergence.
- Momentum: Momentum influences how much the previous weight updates affect the current update.
  - Typical values: [0.0, 0.2, 0.4, 0.6, 0.8, 0.9].
  - Effect: Momentum helps smooth the weight updates by incorporating past gradients, which speeds up convergence and reduces oscillations in the weight updates.

5. Weight Initialization

- Definition: Weight initialization determines the initial values of the weights before training begins. This is an important factor in ensuring proper training convergence.

6. Activation Functions

- Definition: Activation functions introduce non-linearity to the network, allowing it to learn complex patterns.
- Common Functions:
  1. ReLU: Often used in hidden layers to avoid vanishing gradients.
  2. Sigmoid/Softmax: Typically used in output layers for binary/multi-class classification.
  3. Tanh: Historically used in earlier models, but less common now.
  4. Leaky ReLU, ELU, SeLU, Softplus: Variants of ReLU to address issues like dying neurons.

7. Dropout Regularization

- Definition: Dropout is a regularization technique used to prevent overfitting by randomly setting a fraction of the input units to zero at each update during training.

8. Number of Neurons in Each Hidden Layer

- Definition: This parameter refers to the number of neurons in each hidden layer of the network.

- Effect: More neurons can lead to higher capacity for the network to learn complex patterns, but they can also increase computational cost and the risk of overfitting. Optimizing the number of neurons, in combination with the batch size and number of epochs, is critical for obtaining the best results. Each of these hyper parameters plays a crucial role in determining the performance of the CNN model. By carefully tuning them, we can optimize the model for classifying advertisement images, ensuring that it generalizes well on unseen data while avoiding issues like overfitting or slow convergence.

3.2. Model Building

For building the proposed CNN model, Keras is used as the high-level framework on top of the Tensor Flow backend. Keras is a popular open-source neural network library designed to simplify the process of building and training deep learning models. It offers an intuitive and user-friendly API that allows researchers and practitioners to quickly prototype, experiment, and deploy machine learning models. Tensor Flow on the other hand, serves as the low-level library and computational backend that powers Keras. Tensor Flow is an open-source machine learning platform developed by Google, and it is highly optimized for deep learning tasks, particularly in training neural networks efficiently.

By utilizing Keras with TensorFlow, the model can benefit from TensorFlow's scalability and performance while maintaining the simplicity and ease of use provided by Keras. Keras abstracts much of the complexity involved in defining and training models, allowing for rapid experimentation and testing of different CNN architectures.

3.3. Dataset

Since there is a lack of standard datasets specifically focused on advertisements from newspapers, a custom dataset was created for this research. The dataset was compiled from online Malayalam newspapers from january 2023 to December 2023. The newspapers used in the dataset include:

- Malayala Manorama
- Mathrubhumi

To ensure a balanced representation of categories, a dataset was created containing 4800 advertisement images in total. These images are equally distributed across six categories, with 800 images per category. The six categories are as follows

1. Category 1: Job
2. Category 2: Education
3. Category 3: Trade
4. Category 4: Service
5. Category 5: Politics &Administration
6. Category 6: Finance

This dataset allows the CNN model to train on a diverse set of images and ensures that the model can learn to classify advertisements into meaningful categories. The process of creating this custom dataset was essential because no preexisting dataset met the specific needs of this research, such as a comprehensive set of advertisements from multiple English-language newspapers.

Advertisement dataset (4800) is divided in to Training Data set (80% i.e. 3840) and Test Data set (20% i.e. 960).

Training data set is further divided in data for training use (80% of training data) and data for validation use (20% of training data).

3.4. Performance Measures

To compare the results of various models, different performance measures are used including:

- Precision: No. of True Positives / (No. of True positives + No. of False Positives)
- Recall: No. of True Positives / (No. of True Positives + No. of False Negatives)
- F1-score: 2* (Recall * Precision) / (Recall + Precision)
- Accuracy
- Confusion matrix

4. Results

There are three CNN-based architectures utilized as classification models because CNN is thought to be the state-of-the-art in image classification. The outcomes of three CNN-based models that were tested using a dataset of advertisements from newspapers are shown below.

Three models are employed. A basic CNN model makes up Model 1. 'Transfer learning' is used in Model 2. Using a previously trained model with a fresh dataset to train a CNN model quickly and efficiently is called "transfer learning." For transfer learning, a pre-trained ResNet50 model (trained on the Image Net dataset) is employed. Model 3 employs 'Transfer learning' as well as a pre-trained ResNet50 model, and a few extra layers are added at the very end of the model architecture.

S.No.	Model	Accuracy
1.	Model1(Simple CNN)	65%
2.	Model2(ResNet50 + Classifier)	68%
3.	Model3(Finetuned-ResNet50)	74%

Table 1: Summary of the results obtained from the evaluated models

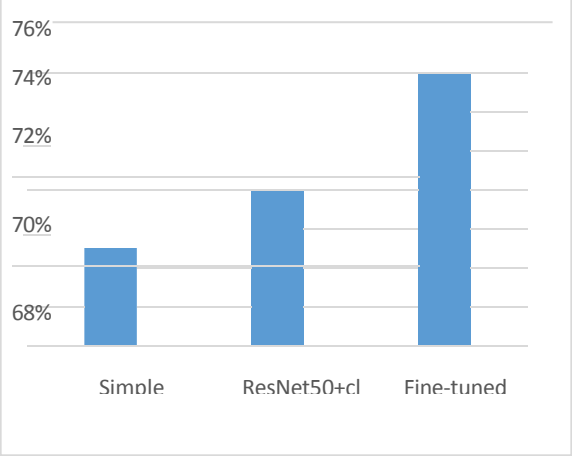

Fig 3. Comparison of evaluated models

5. Conclusion

The newspaper advertisement classification and its analysis makes it possible how each leading newspaper daily contribute towards advertising industry. Based on such classification and analysis the statistics behind advertisement

can be revealed. It has significant influence in the market. The people who are interested with a particular category of product can choose a specific newspaper which promote the advertisements of that products highly. The products and its supplementary items can focus on that particular daily which will attract more customers to a single point. Definitely there will be a ranking of newspapers according to the types and count of the advertisements they promote. Also a statistical analysis of increase or decrease of products advertisements gives a picture of industrial growth of each specific items which will be helpful for business class people and entrepreneurs to make wise decisions.

6. References

[1] Peleato, R. A., Chappelier, J. C., & Rajman, M. (2000). Using information extraction to classify newspapers advertisements. In Proceedings of the 5th International Conference on the Statistical Analysis of Textual Data, Lausanne, Switzerland (pp. 28-30).

[2] Chu, W. T., & Chang, H. Y. (2016). Advertisement Detection, Segmentation, and Classification for Newspaper Images and Website Snapshots. In 2016 International Computer Symposium (ICS) (pp.396-401). IEEE https://doi.org/10.1109/ICS.2016.0086

[3] Vo, A. T., Tran, H. S., & Le, T. H. (2017). Advertisement image classification using convolutional neural network. In 2017 9th International Conference on Knowledge and Systems Engineering (KSE) (pp. 197-202). IEEE. https://doi.org/10.1109/KSE.2017.811945

[4] Gatos, S.Mantzariz ,K.Chandrinos, "Integrated algorithms for newspaper page decomposition and article tracking"Proceedings of international conference on document analysis and recognition",pp 559-562,1999

[5] F.Liu, Y.Luo, M.Yoshilkawa" A new component based algorithm for newspaper layout analysis" Proceedings of international conference on Document analysis and Recognition,pp 1176-1180,2001

[6] Mitchell and H.Yan " Newspaper document analysis featuring connected line segmentation" Proceedings of International conference on Document Analysis and Recognition pp 1181-1185,2001

[7] A.Bansal, S.Choudhary, S.Roy and J.B Srivasthava" Newspaper article extraction using hierarchical fixed point model". Proceedings of IAPR workshop on Document analysis systems,2014

[8] F.Fouzi,, J.L Hong and M.Belkhatir "web page segmentation for extracting images and their surrounding contextual information"Proceedings of ACM international Conference on Multimedia, pp 649-652,2009

[9] Khaled Almgren, Murali Krishnan , Fatima Aljanobi and Jeongkyu Lee "A Deep Learning Approach to Detect Advertisements from Magazines" Entropy ,Published: 17 December 2018

[10] Pooja Jain1 , Kavita Taneja2 , Harmunish Taneja "Convolutional Neural Network Based Advertisement Classification Models for Online English Newspapers", Turkish Journal of Computer and Mathematics Education, Vol.12 No.2 (2021), 1687- 1698

[11] Rene Y. Choi; Aaron S. Coyner; Jayashree Kalpathy-Cramer; Michael F. Chiang; "Introduction Machine Learning, Neural Networks, and Deep Learning" Translational Vision Science & Technology February 2020, Vol.9, 14

[12] LeCun, Y., Kavukcuoglu, K., & Farabet, C. (2010). Convolutional networks and applications in vision. Proceedings of 2010 IEEE International Symposium on Circuits and Systems, 253–256. https://doi.org/10.1109/ISCAS.2010.5537907

[13] Torrey, L., & Shavlik, J. (2010). Transfer learning. In Handbook of research on machine learning applications and trends: algorithms, methods, and techniques (pp. 242-264). IGI global.

[14] Nibeesh K, Sreejith C M.Tech Computational Linguistics Dept of Computer Science and Engineering GovernmentEngineering College Sreekrishnapuram, "Malayalam_Text_Classification_for_Efficient_News_Filtering_using_Support_Vector Machine" https://www.academia.edu/6276563

[15] P.A.Charde, S.D.Lokhand, "Classification Using K Nearest Neighbor for Brain Image Retrieval" International Journal of Scientific & Engineering Research, Volume 4, Issue 8, August-2013, ISSN 2229-5518

[16] Kun Fang ," Naive Bayes Image Classification Based on Multiple Feature", Computer Software and Media Applications (2019) Volume 2, doi: 10.24294/csma.v2i1.1171

[17] Le Hoang Thai, Tran Son Hai "Image Classification using Support Vector Machine and Artificial Neural Network" I.J. Information Technology and Computer Science, 2012, 5, 32-38 Published Online May 2012 in MECS (http://www.mecs-press.org/) DOI: 10.5815/ijitcs.2012.05.05

[18] Pin Wang , En Fan , Peng Wang," Comparative Analysis of Image Classification Algorithms Based on Traditional Machine Learning and Deep Learning" Pattern Recognition Letters (2020), doi: https://doi.org/10.1016/j.patrec.2020.07.

[19] Shrey Srivastava , Amit Vishvas Divekar, Chandu Anilkumar, Ishika Naik, Ved Kulkarni and V. Pattabiraman "Comparative analysis of deep learning image detection algorithms" *Journal of Big Data* **volume 8**, Article number: 66 (2021)

[20] Boxuan Zhong , Qian Ge, Bhargav Kanakiya, Ritayan Mitra , Thomas Marchitto and Edgar Lobaton, Institute of Arctic and Alpine Research University of Colorado Boulder," A Comparative Study of Image Classification Algorithms for Foraminifera Identification"

[21] Mahmoud Smaida , Dr. Yaroshchak Serhii ,"Comparative Study of Image Classification Algorithms for Eyes Diseases Diagnostic", International Journal of Innovative Science and Research Technology ISSN No:-2456-2165, Volume 4, Issue 12, December – 2019

[22] Mingyuan Xin and Yong Wang," Research on image classification model based on deep convolution neural network" EURASIP Journal on Image and Video Processing (2019) 2019:40, https://doi.org/10.1186/s13640-019-0417-8