

Convolutional Neural Network Based Advertisement Classification Models for Online English Newspapers

Pooja Jain¹, Kavita Taneja², Harmunish Taneja³

^{1,2}Panjab University, Department of Computer Science & Application, Chandigarh/India

³DAV college, Department of Computer Science & Information Tech., Chandigarh/India

Article History: Received: 10 November 2020; Revised: 12 January 2021; Accepted: 27 January 2021; Published online: 05 April 2021

Abstract: Image processing for knowledge management and effective information extraction is the key element for steering towards society 5.0. There has been a substantial research and progress in the area of image recognition and classification in the recent years but at the same time, there is a lack of significant work in the field of advertisement image classification from online English newspapers. This research paper analyses and compares various popular image classification techniques to find out the most suitable technique for advertisement image classification problem. Automatic feature extraction without any prior knowledge of features makes Convolutional Neural Networks (CNN) the most suitable technique for advertisement image classification. This paper further explores and implements three different CNN-based image classification models that can classify advertisement images from online English newspapers into four pre-defined categories including Admission-notices, Job-advertisements, Sales and Promotional advertisements and Tenders. These models are trained and tested on an advertisement image dataset collected from four different online English newspapers over a time frame of 15 months. Fine-tuned ResNet50 Model using ‘Transfer-learning’ is found to be the most suitable model for this advertisement image classification task with results exhibiting around 74% accuracy. This CNN-model based automated classification of advertisement images will help newspaper readers in performing exhaustive advertisement search in a category of their own interest, saving the time and efforts of sequential manual search across a range of multiple newspapers. Also, the proposed research will help in performing advertisement analysis and studies.

Keywords: Advertisement image classification, Convolutional Neural Networks (CNN), Image classification techniques, Residual Networks (ResNet), Transfer learning

1. Introduction

Online newspapers are in much trend these days for the convenience of accessing information on our laptops, mobiles, smartphones, tablets and desktops etc. anywhere at any time. Moreover, home locked situations during unprecedented circumstances like pandemics which restrict the access to the printed copy of newspapers have further boosted this trend many folds. With younger population increasingly embracing technology, this trend of online newspaper reading is going to last for many years to come. Along with the news articles, advertisements in the newspapers are of much interest. Government departments, recruitment agencies, educational institutes, private companies etc. use newspaper advertisements as a primary source for advertising tenders, jobs, admission notices, sales and promotions etc. and people anxiously wait for these advertisements to be out in the newspaper. Students may be interested in admission-notices whereas job aspirants may look for job-advertisements in the newspapers. A contractor may be interested in the relevant tender-notice and a shopping enthusiast may be looking for sales and promotional advertisements. But online newspapers do not give this type of category-wise personalised search options. Also, no search engine including Google has a primary purpose of searching advertisements from online newspapers. As a result, when we search for some advertisement in the newspapers through search portals, we may get hundreds of images but needless to say that only a few of them are relevant. Hence the reader is left with no option but to sequentially go through all the newspapers and manually search the relevant advertisements for himself or herself. This sequential manual search is very time consuming and tedious specially when the reader is searching for a particular advertisement across a range of newspapers. An advertisement image classification model which can classify each input advertisement into various pre-defined advertisement categories can be very helpful in performing this type of personalised advertisement search. When combined with OCR (Optical Character Recognition) techniques and user-friendly search interface, this advertisement image classification model can help a reader in performing category-wise advertisement search across a range of newspapers saving the time and effort of sequential manual search.

Advertisement image classification is typically a supervised machine learning problem which involves two phases. The first phase is the learning or training phase in which a classification model is created using a classification technique (learning algorithm) and the model is trained on advertisement dataset. Second phase is the recognition or classification phase where the advertisement classification model is used to classify the new advertisement images into different pre-defined categories. Many image classification techniques are available to

choose from and each of the technique has its own advantages and disadvantages. SVM is one of the most popular image-classification techniques used in several applications including facial expression classification, bioinformatics (classification of genes, protein remote homology detection, cancer classification etc.), Generalized Predictive Control (GPC), text and hypertext classification and many more. On the other hand, CNN is also used in many real-world applications such as face recognition (in social media applications), image analysis (in health care), OCR, object-detection (for driver-less cars) etc. and has become the first choice for image recognition and classification tasks among many researchers. The proposed research explores various popular image classification techniques including K-NN, Decision trees, Naïve Bayes, SVM and CNN and chooses the most suitable image classification technique for advertisement image classification from online English newspapers. The chosen technique is employed in three different models and these models are evaluated on the advertisement dataset and their validity is ascertained.

The major contributions of the proposed research are: (1) Choosing the most relevant image classification technique for advertisement image classification from English newspapers, (2) Creation of advertisement dataset from online English newspapers, (3) Designing, training & testing three different advertisement image classification models using the advertisement dataset created.

The rest of the paper is organized as follows: Popular image classification techniques are introduced and a comparative analysis is drawn and the most suitable classification technique is chosen for advertisement image classification from online English newspapers. Related work in the field of advertisement image classification is presented in the next section. Following section presents the proposed CNN-based advertisement image classification model building and the advertisement dataset used followed by 'Result and Discussion' section that elaborates the implemented CNN models along with their performance analysis. Finally, the conclusion section concludes the paper.

2. Popular Image Classification Techniques

Popular image classification techniques for supervised learning are presented as follows:

2.1. K-NN (k-nearest neighbor)

K-NN(Cover & Hart, 1967) is one of the simplest supervised machine learning algorithm for image classification. Training phase is very simple and involves storing feature vectors and labels of the training image-set. In the recognition phase, K-NN first locates the k-nearest data samples to the query data sample and assigns a class label which is most frequent in the k-nearest data samples i.e. it classifies the unknown data point on the basis of its closest neighbours whose classes are already known. KNN algorithm considers all the features equal in similarity computation leading to misclassification when only a small subset of features are useful for classification purposes (Kim¹ et al., 2012).

2.2. Decision Trees

Decision trees(Murthy, 1998) classify query samples based on their sorted feature values. In a decision tree, each node represents a feature that a data-sample may possess and each branch represents a value that a particular feature node can assume. Leaves represent the final decisions (classes in which the instances are finally classified). Classification process starts at root node. Different branches are followed based on the feature values present in the query sample and finally a leaf node is reached which is the result of the classification algorithm. The well-known algorithms for building decision trees are ID3, C4.5, EC4.5, Rainforest and PUBLIC etc. (Kotsiantis, 2007).

2.3. Naïve Bayes

Based on the Bayes Theorem, Naïve Bayes (Rish, 2001; Lewis, 1998) is a statistical learning technique that predicts membership probabilities (probability of a data sample or record belonging to a specific class) for each class based on the feature vector. The class whose probability is the highest is considered as the most likely class. The name naive is used because it assumes the features that go into the model is independent of each other. This assumption of independence of features rarely holds true in the real world and hence Naïve Bayes classifiers are generally less accurate as compared to other more complex learning algorithms (Kotsiantis, 2007).

2.4. SVM

Support Vector Machine (SVM) (Cortes & Vapnik, 1995) represents the data samples as points in space and finds out the optimal hyperplane which separates the data samples belonging to two different classes with maximum margin. The test or query data sample is also mapped onto that same space and its class is predicted based on which side of the hyperplane it falls on. Training and testing speed of SVMs is inherently very slow. Also, the selection of the kernel function parameters for SVM is not straight forward (Kim¹ et al., 2012). SVM classification requires useful features which need to be extracted from the entire set of training instances using techniques like K-Means, Genetic Algorithm, F-score, SVM-RFE (Recursive Feature Elimination) and ReliefF etc. (Kaur & Kalra, 2016).

2.5. CNN

CNN (LeCun et al., 2010; Krizhevsky et al., 2017; Shaheen & Verma, 2016) is a deep learning architecture that contains many hidden layers including convolutional layer, pooling layers, fully connected layers and flatten layers. Convolutional layers detect various low-level and high-level features by applying different filters to the sample image. A nonlinear activation function (generally Rectified Linear Units (ReLU)) is applied to the features obtained after convolution. After each convolutional layer, a pooling layer is added. These convolutions + pooling blocks can be repeated many times. The last pooling layer is followed by flattening layer. Next step is to add a fully connected layer. One or more fully connected layers may follow and finally, classification is performed using different classification techniques (for e.g. Softmax) and classification output is obtained. Three main properties of CNN namely : ‘Sparse connectivity’, ‘Shared weights’ and ‘Pooling’ work towards dimensionality reduction of the network and reduce the training time to a great extent (Hasan et al., 2019). Training time can further be decreased using Graphical Processing Units (GPUs).

2.6. Comparative Analysis of Popular Image Classification Techniques

The important features and limitations of different classification techniques are summarized in Table 1.

Table 1. Comparison of popular image classification techniques

S.No.	Classification Technique	Features	Limitations
1.	K-NN	<ul style="list-style-type: none"> • Zero cost of learning process • Only one parameter (k) needs to be tuned. • Classes need not be linearly separable • Well suited for multimodal classes. 	<ul style="list-style-type: none"> • Large storage requirements (for both training and classification) • Sensitive to the similarity function chosen for comparison. • Difficult to choose suitable ‘k’ • Extremely computation intensive • Not suitable for high dimensional feature space. • Not robust to missing values • Very sensitive to irrelevant features and intolerant to noise
2.	Decision Trees	<ul style="list-style-type: none"> • Comprehensibility: one can quickly understand why a data sample is classified as belonging to a particular class • Univariate decision trees are quite fast • Resistant to noise. 	<ul style="list-style-type: none"> • Instability: A minor change in the data may cause significant change in the decision tree. • Computationally intensive. • Space and time complexities are high • Not suitable for regression analysis
3.	Naïve Bayes	<ul style="list-style-type: none"> • Less training-time • Less space requirements for both training and classification • Robust to missing values 	<ul style="list-style-type: none"> • ‘Zero Conditional Probability’ Problem • Less accurate because of the strong feature independence assumption
4.	SVM	<ul style="list-style-type: none"> • Unaffected by large no. of features: suits well when number of features are large in comparison to the number of training samples. • Training optimization: Always attains a global minimum (not local). 	<ul style="list-style-type: none"> • Does not support automatic feature extraction • Low speed of the training. • Discrete data presents problem • Inherently a binary classification approach • Lot of model parameters to tune.

5.	CNN	<ul style="list-style-type: none"> • Automatic feature selection and extraction • Does not require prior knowledge of features. • Classifier takes very less space in memory and can classify query samples in very less time compared to training time. 	<ul style="list-style-type: none"> • May suffer from overfitting • Needs very big dataset for training • Computationally intensive training • Difficult to find the efficient network architecture • Black-box • Not robust to missing values • Lot of model parameters to tune • Sensitive to irrelevant features • May suffer from overfitting
----	-----	---	---

2.7. Classification Technique for Advertisement Images

As discussed in the previous section, many image classification techniques are available to choose from. Being simple and easy to understand, K-NN is a basic classification approach and gives accepted results but when it comes to advertisement image classification, no. of dimensions of the feature space is too large to compute pixel by pixel feature space distances making K-NN unsuitable for advertisement image classification. The information to be extracted from online newspapers / web contents can be structured, unstructured and semi-structured (Dhuria et al., 2016). Decision Trees are the most comprehensible image classification technique but creating a decision tree which includes all the features required to classify the advertisement images in different categories would be computationally very-very intensive and hence not feasible. Strong ‘feature independence assumption’ makes Naïve Bayes classification technique unfit for advertisement image classification. SVM is a commonly used technique for image classification but the major limitation of SVM is that it does not support automatic feature extraction and various other techniques are required to extract useful features from the entire set of features prior to the classification process. This generally needs prior knowledge of features involving domain experts and the datasets need to be prepared according to the feature set identified. Also, SVM is basically a binary classification approach. For multi-class classification problems like advertisement image classification, the classification task needs to be converted in to a set of multiple binary classification problems adding computational complexities as the no. of classes increase. On the other hand, CNN is a multi-class classification technique that just needs an image dataset for training and automatically learns the features from the dataset. This automatic feature extraction property makes CNN the most suitable technique for advertisement image classification from online English newspapers where only images are available with the class labels to learn from, without any prior knowledge of the features.

3. Related Work in Advertisement Image Classification

There has been tremendous research efforts in object image recognition and classification field (Mikolajczyk et al., 2005; Li & Fei-Fei, 2007; Frome et al., 2007; LeCun et al., 2010; Krizhevsky et al., 2017) but at the same time there is a lack of significant work explicitly in the field of recognition and classification of advertisement images.

Peleato et al. (2000) classified newspaper advertisements into four different categories including real estate, vehicles, employment and others by combining Naïve Bayes classifiers and information extraction techniques. This model suffers from a basic problem that it needs textual content of the advertisements for classification but newspaper advertisements are images and are not directly available in text format.

Chu & Chang (2016) suggested a framework for advertisement analysis in newspaper images and website snapshots. Newspaper images and website snapshots are segmented to extract advertisement candidates followed by their classification using rule-based filters (advertisement size and placement in the front page of China Times) and learning-based filters (CNN features + SVM classification). Rule-based filters may not be a good choice when it comes to advertisement classification across multiple newspapers as different newspapers may have different rules for placing the advertisements. Also, this market is totally customer driven where rules may change according to the customer’s need and requirements.

Another CNN-based advertisement-image classification model (nLmF-CNN) was proposed by Vo et al. (2017) which identifies whether an online advertisement is displayed clearly or not. This research finds out the suitable values of no. of conv layers (n) and no. of filters (m) for a dataset of online advertisements on website snapshots. Identification of suitable hyperparameters for other datasets including newspaper advertisements dataset remains an open challenge. Also, the proposed model is a binary classification model which classifies the instances into two classes. Multi-class classification of advertisement images is not addressed here.

4. Proposed CNN Model for Classification of Advertisement Images

Since 'Imagenet-classification' in 2012 (Krizhevsky et al., 2017), CNN is considered as state-of-the-art for image recognition and classification. The proposed research explores various CNN-based image classification models for classifying the advertisement images collected across a range of newspapers.

4.1. CNN- Hyperparameters

Many CNN parameters need to be tuned for achieving the desired accuracy including:

Batch-size: In CNN training, batch size is the number of training samples shown to the network before the weight updation. CNNs are sensitive to batch size.

No. of Epochs: The number of epochs represents the total number of times the entire training data samples are presented to the network for training purposes.

Optimization Algorithm: Optimization algorithms reduce the error function by changing the attributes (such as weights and learning rate) to provide the most accurate results possible. Different optimization algorithms (Vani & Rao, 2019) are used in neural networks including: SGD, Adadelta, RMSprop, Adam, AdaMax, Nadam, Adagrad etc. It is desired to choose an optimization algorithm that gives the best accuracy.

Learning Rate and Momentum: Weight updation at the end of each batch is controlled by the learning rate whereas the momentum governs how much the previous update impact the current weight update. Typical values of learning rate can be: [0.001, 0.01, 0.1, 0.2, 0.3] etc. and momentum is usually chosen from [0.0, 0.2, 0.4, 0.6, 0.8, 0.9] etc.

Weight Initialization: There are different weight initialization techniques including 'uniform', 'normal', 'Random_uniform', 'Random_normal', 'zeros', 'ones', 'glorot_normal', 'glorot_uniform', TruncatedNormal, 'lecun_uniform', Identity, Orthogonal, Constant, VarianceScaling, 'he_normal', 'he_uniform' etc. Different Weight Initialization techniques can be used for different layers according to the activation function used in each layer (Li et al., 2020).

Activation functions: The non-linearity of individual neurons is controlled by the activation function. It controls when a neuron will fire. Many activation functions are available including: Linear, ReLU, Tanh, Sigmoid, Hard_sigmoid, Softmax, Softplus, Softsign, SeLU, ELU & Exponential etc. In most of the deep learning applications, Softmax function or Sigmoid function are mostly used in the output layer while the ReLU units are used in the hidden layers (Nwankpa et al., 2018).

Dropout Regularization: Dropout is a technique to avoid overfitting (when a deep learning model starts cramming the training samples). Dropout rate can vary from 0.0 to 0.9 (Srivastava et al., 2014).

Number of neurons in each hidden layer: It is an important parameter to tune. It should ideally be optimized with batch size and number of epochs.

4.2. Model Building

Keras (Gulli & Pal, 2017)(open source neural network library) is used on top of Tensorflow (Abadi et al., 2016) backend to implement various CNN models in Python programming language. TensorFlow is an open-source platform for machine learning. It was primarily developed by Google for Deep neural network research. On the other hand, Keras is a high-level API (Application Programming Interface) that work as a wrapper to low-level libraries of Tensorflow.

4.3. Dataset

There is a lack of standard data set of advertisements from English newspapers and hence the authors have created their own dataset of advertisements from various online English newspapers which were free to download at the time of advertisement data collection (May, 2019 to Sept 2020) including 'Times of India', 'Hindustan Times', 'Indian Express' and 'The Tribune'. A balanced dataset of 4400 images is created with 1100 images in each of the four categories namely:

- (1) Admission_Notices (including admission-notices, advertisements of educational institutes, coaching classes, scholarships etc.)
- (2) Job_Ads
- (3) Sales_and_Promotion
- (4) Tenders (including tenders, bids, auctions, request for proposals etc.)

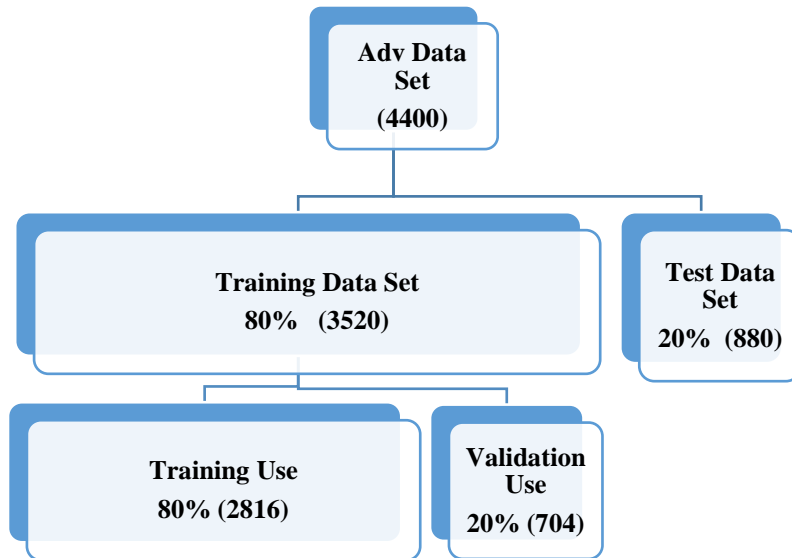


Figure 1. Division of dataset in Training, Validation and Test sets

As shown in Figure 1, Advertisement dataset (4400) is divided into Training Data set (80% i.e. 3520) and Test Data set (20% i.e. 880). Training data set is further divided into data for training use (80% of training data i.e. 2816) and data for validation use (20% of training data i.e. 704).

4.4. Performance Measures

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

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

5. Results and Discussion

Three CNN models with different hyperparameters are trained on the advertisement dataset and are evaluated to ascertain the validity of each model.

5.1. Model 1 (Simple CNN model)

A simple CNN model (Figure 2) with first layer as convolution layer with 32 filters and 3x3 kernel size + 'ReLU' activation + 'MaxPooling' with (2x2) kernel size + 'Dropout' 0.5 or 50% followed by another convolution layer with 64 filters and 3x3 kernel size + 'ReLU' + 'MaxPooling' with (2x2) kernel size + 'Dropout' 0.5 or 50%. These layers are followed by one fully connected layer with 128 output units + 'ReLU' activation + 'Dropout' 0.5 or 50% and finally a fully connected layer with 4 output units + 'Softmax' classification which classifies the input image into one of the four pre-defined categories (Admission_Notice/ Job_Ads/ Sales_and_promotions / Tenders).

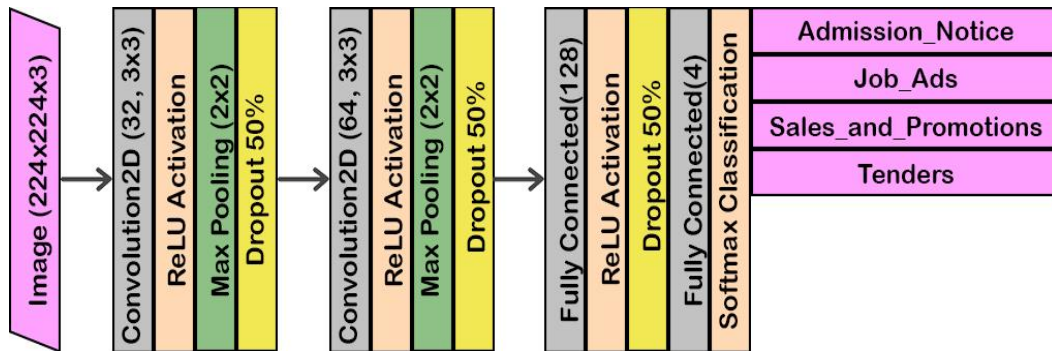


Figure 2. Model 1 (Simple CNN) architecture

Model 1 is trained using ‘Categorical_crossentropy’ loss function along with ‘Adam’ optimizer and a batch-size of 32. Training is done for 100 epochs. Accuracy on the test set (880 images) comes out to be 0.65 (65%). Table 2 shows the evaluation results of Model 1 on the test set. Figure 3a shows the ‘Confusion matrix’ and Figure 3b presents the training and validation accuracy curves.

Table 2. Test results from Model 1

	Precision	Recall	F1-score	Support
Admission_Notice	0.65	0.50	0.56	223
Job_Ads	0.57	0.50	0.53	216
Sales_And_Promotion	0.76	0.86	0.80	228
Tenders	0.61	0.75	0.68	213

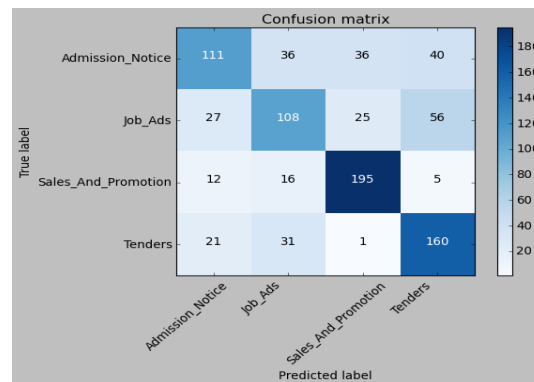


Figure 3a. Confusion matrix (Model 1)

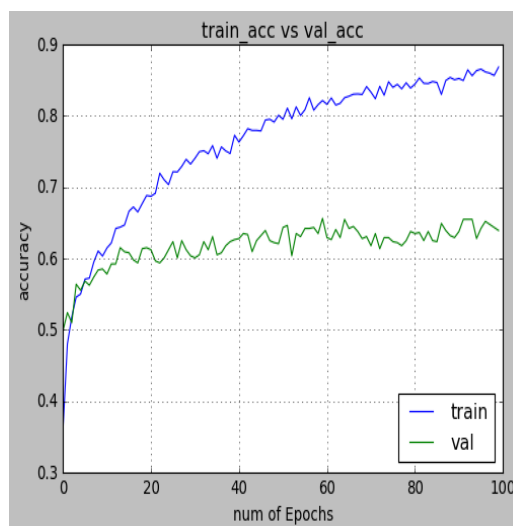


Figure 3b. Accuracy curves (Model 1)

Figure 3a shows that Model 1 recognises ‘Sales_and_Promotion’ advertisements better (195 correctly classified out of total 228 ‘Sales_and_Promotion’ advertisements whereas 33 are misclassified) than the other categories followed by ‘Tenders’ (160/ 213), ‘Admission_Notice’ (111/ 223) and ‘Job_Ads’ (108/ 216).

CNN requires huge datasets in order to achieve higher accuracies. Since, our dataset is limited to 4400 images, the achieved accuracy (65%) is low as expected. It is also observed that when more layers are added to this simple network, the performance of the network further degrades resulting in lesser accuracies.

5.2. Model 2 (ResNet50 + Classifier)

In order to achieve higher accuracies with small dataset, Model 2 uses ‘Transfer learning’. ‘Transfer learning’ (Torrey & Shavlik, 2010; Pan & Yang, 2010) is an easy way to train a CNN model with small dataset and short computation and training time with the help of a pre-trained model and the new dataset. Here, pretrained ResNet50 model (trained on ImageNet dataset) is used for Transfer learning. Residual networks (ResNet) (He et al., 2016) use a technique called ‘skip connections’ which helps in training very deep neural networks without facing vanishing/exploding gradient problem. Model 2 uses ResNet50 for feature extraction. As shown in Figure 4, only the classifier is modified and trained to classify advertisement images keeping the weights of all the above layers unchanged.

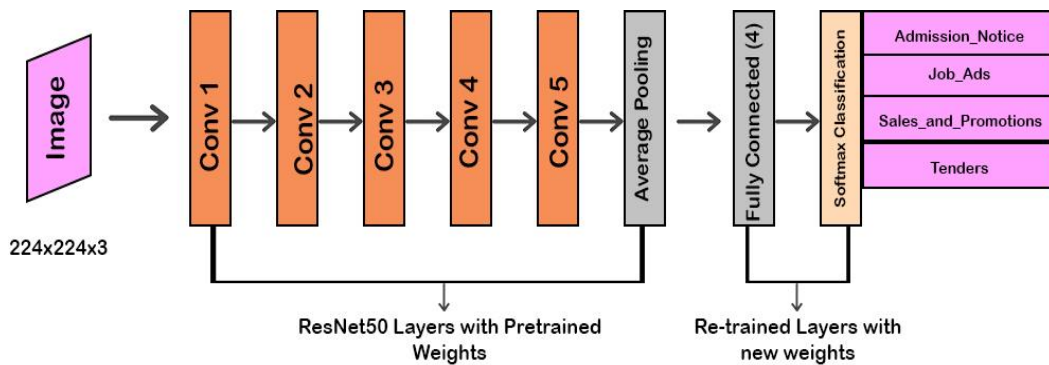


Figure 4. Model 2 (ResNet + classifier) architecture

Model 2 also uses ‘Categorical_crossentropy’ loss function along with ‘Adam’ optimizer and a batch-size of 32. Training is done for 100 epochs. Model Accuracy of 0.68 (68%) is achieved on the Test set (880 images). Evaluation results of Model 2 on the Test set are presented in Table 3 along with ‘Confusion matrix’ (Figure 5a) and training and validation accuracy curves (Figure 5b).

Table 3. Test results from Model 2 (ResNet50 + classifier)

	Precision	Recall	F1-score	Support
Admission_Notice	0.64	0.56	0.60	223
Job_Ads	0.57	0.61	0.59	216
Sales_And_Promotion	0.87	0.78	0.82	228
Tenders	0.66	0.77	0.71	213

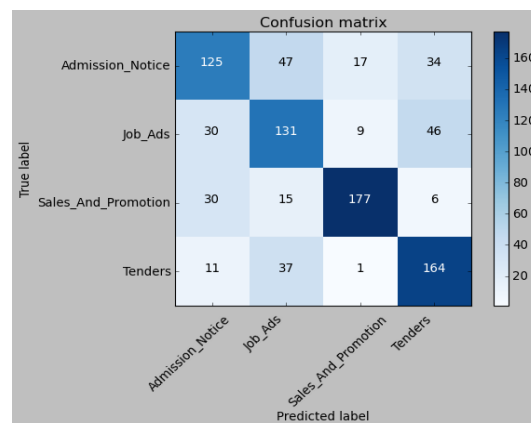


Figure 5a. Confusion matrix (Model 2)

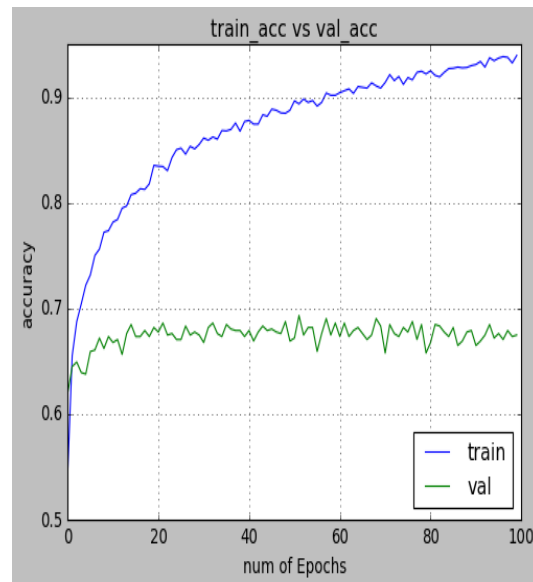


Figure 5b. Accuracy curves (Model 2)

The confusion matrix for Model 2 (Figure 5a) shows the improvement in the recognition of ‘Tenders’ (164/213), ‘Job_Ads’ (131/216) and ‘Admission_Notice’ (125/223) as compared to Model 1 but the recognition accuracy of ‘Sales_and_Promotion’ advertisements is decreased (177/ 228) as compared to Model 1 (195/228). Overall accuracy is improved (68%) as reflected in the validation accuracy curves (Figure 5b).

5.3. Model 3 (Fine-tuned ResNet50)

Model 3 also uses ‘Transfer learning’ with pretrained ResNet50 model but this time few more layers at the end of the model architecture (Global_Average_Pooling2D + Fully connected (fc-1) with 512 output units + ‘ReLU’ + Dropout (0.5 or 50%) + Fully connected layer (fc-2) with 256 output units + ‘ReLU’ + Dropout (0.5 or 50%) + Fully connected layer (fc-3) with 4 output units + ‘Softmax’ classification layer) are trained using advertisement dataset along with the classifier as shown in Figure 6.

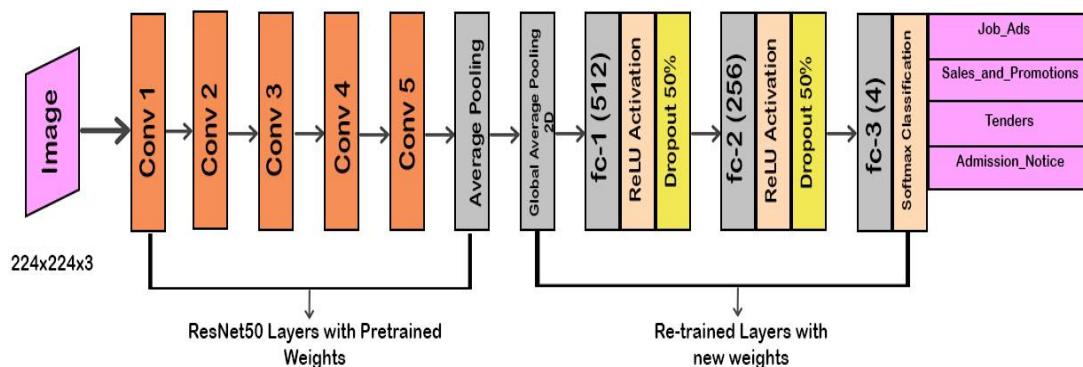


Figure 6. Model 3 (Finetuned ResNet50) architecture

Similar to Model 1 and Model 2, Model 3 also uses ‘Categorical_crossentropy’ loss function along with ‘Adam’ optimizer and a batch-size of 32 is used for training for 100 epochs. Here, model accuracy of 0.74 (74%) is achieved on the test set (880 images). Table 4 shows the evaluation results of Model 3 on the test set and Figure 7a shows the ‘Confusion matrix’ whereas Figure 7b shows the training and validation accuracy curves.

Table 4. Test results from Model 3 (Fine-Tuned ResNet50)

	Precision	Recall	F1-score	Support
Admission_Notice	0.72	0.64	0.67	223
Job_Ads	0.62	0.68	0.65	216
Sales_And_Promotion	0.89	0.84	0.86	228
Tenders	0.72	0.78	0.75	213

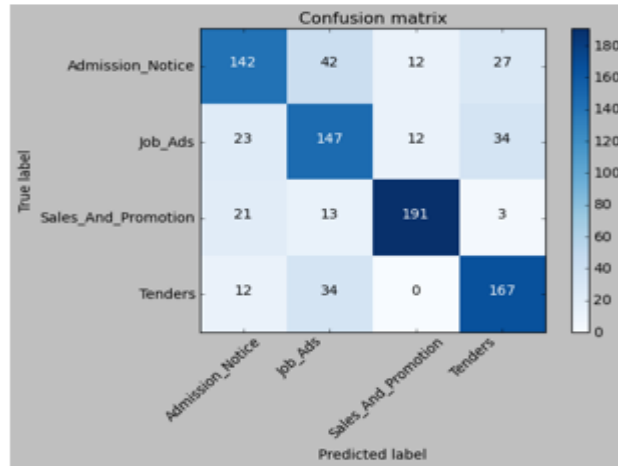


Figure 7a. Confusion matrix (Model 3)

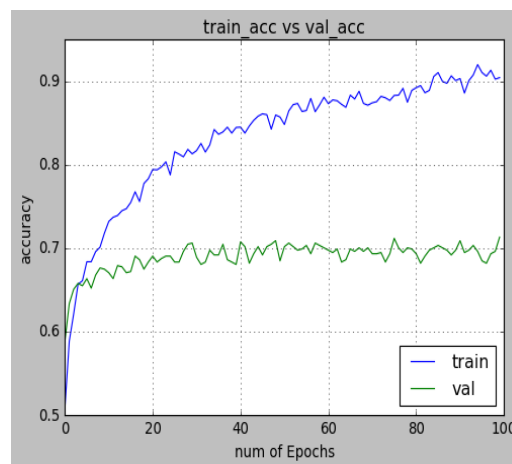


Figure 7b. Accuracy curves (Model 3)

The confusion matrix for Model 3 (Figure 7a) shows the improvement in the recognition of all the advertisement categories including ‘Sales_And_Promotion’ (191/228), ‘Tenders’ (167/ 213), ‘Job_Ads’ (147/216) and ‘Admission_Notice’ (142/223) as compared to Model 2 and hence the overall accuracy is also improved (74%) significantly as reflected in the validation accuracy curves (Figure 7b).

6. Comparison of the Results obtained

The results of all the three CNN-based models trained and tested on advertisement dataset from online English Newspapers are summarized in Table 5. It is found that Finetuned-ResNet50 which is trained on last few layers gives the best accuracy of 74% among the three evaluated models and is the most suitable model for advertisement image classification from online English newspapers.

Table 5. Summary of the results obtained from the evaluated models

S.No.	Model	Accuracy
1.	Simple CNN-Model	65%
2.	ResNet50 + Classifier	68%
3.	Finetuned-ResNet50	74%

7. Conclusion

The choice of image classification technique (learning algorithm) always depends on the task at hand. CNN just needs an image dataset for training and automatically learns the features from the dataset and this property makes it the most suitable technique for advertisement image classification from online English newspapers. There is a lack of standard dataset of advertisements from English language newspapers and the authors have created their own dataset of advertisements from four different online English newspapers. The proposed

research designed, implemented and trained three CNN-based advertisement image classification models on the advertisement dataset. These models can classify the input advertisements from different English newspapers into four pre-defined advertisement categories. Simple CNN model with two convolutional layers gives less accuracy of 65% as CNN requires huge dataset for training but the training dataset of advertisements is small in size (4400 images). Using ResNet50 for feature extraction and training the classifier on the advertisement dataset improves the accuracy (68%). Fine-tuned ResNet50 model (which is trained on last few layers using advertisement dataset) is found to be the most suitable model exhibiting around 74% accuracy. Future work includes increasing the accuracy of the model by increasing the size of the advertisement dataset so that more image samples are available to learn from. Future scope also includes filtering out the non-advertisements before advertisement images are classified in to different categories. More advertisement categories like: Public Notices, Matrimonial advertisements, Remembrance messages, Political advertisements etc. can be added in the future extension. The proposed research when clubbed with OCR techniques can also provide for keyword-based advertisement search in different advertisement categories in online English newspapers enhancing reader's online newspaper reading experience for advertisement search many folds.

References

1. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)* (pp. 265-283).
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. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.
4. Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21–27. <https://doi.org/10.1109/TIT.1967.1053964>
5. Dhuria, S., Taneja, H., & Taneja, K. (2016). An optimal approach for extraction of Web Contents using Semantic Web framework. In *Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies* (pp. 1-5).
6. Frome, A., Singer, Y., Sha, F., & Malik, J. (2007). Learning Globally-Consistent Local Distance Functions for Shape-Based Image Retrieval and Classification. *2007 IEEE 11th International Conference on Computer Vision*, 1–8. <https://doi.org/10.1109/ICCV.2007.4408839>
7. Gulli, A., & Pal, S. (2017). *Deep Learning with Keras*. Packt Publishing Ltd.
8. Hasan, M., Ullah, S., Khan, M. J., & Khurshid, K. (2019). COMPARATIVE ANALYSIS OF SVM, ANN AND CNN FOR CLASSIFYING VEGETATION SPECIES USING HYPERSPECTRAL THERMAL INFRARED DATA. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-2/W13*, 1861–1868. <https://doi.org/10.5194/isprs-archives-XLII-2-W13-1861-2019>
9. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
10. Kaur, S., & Kalra, S. (2016). Feature extraction techniques using support vector machines in disease prediction. In *Proceedings of the 4th International Conference on Science, Technology and Management (ICSTM-16), India International Centre, New Delhi*.
11. Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160(1), 3-24.
12. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90. <https://doi.org/10.1145/3065386>
13. 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>
14. Lewis, D. D. (1998). Naive (Bayes) at forty: The independence assumption in information retrieval. In *European conference on machine learning* (pp. 4-15). Springer, Berlin, Heidelberg. <https://doi.org/10.1007/BFb0026666>
15. Li, H., Krček, M., & Perin, G. (2020). A Comparison of Weight Initializers in Deep Learning-Based Side-Channel Analysis. In *International Conference on Applied Cryptography and Network Security* (pp. 126-143). Springer, Cham. <http://eprint.iacr.org/2020/904>

16. Li, L. J., & Fei-Fei, L. (2007). What, where and who? classifying events by scene and object recognition. In *2007 IEEE 11th international conference on computer vision* (pp. 1-8). IEEE. <https://doi.org/10.1109/ICCV.2007.4408872>
17. Mikolajczyk, K., Leibe, B., & Schiele, B. (2005). Local features for object class recognition. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1* (Vol. 2, pp. 1792-1799). IEEE. <https://doi.org/10.1109/ICCV.2005.146>
18. Murthy, S. K. (1998). Automatic construction of decision trees from data: A multi-disciplinary survey. *Data mining and knowledge discovery*, 2(4), 345-389.
19. Nwankpa, C., Ijomah, W., Gachagan, A., & Marshall, S. (2018). Activation Functions: Comparison of trends in Practice and Research for Deep Learning. *ArXiv:1811.03378 [Cs]*. <http://arxiv.org/abs/1811.03378>
20. Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345–1359. <https://doi.org/10.1109/TKDE.2009.191>
21. 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).
22. Rish, I. (2001). An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence* (Vol. 3, No. 22, pp. 41-46).
23. Kim¹, J., Kim, B. S., & Savarese, S. (2012). Comparing image classification methods: K-nearest-neighbor and support-vector-machines. In *Proceedings of the 6th WSEAS international conference on Computer Engineering and Applications, and Proceedings of the 2012 American conference on Applied Mathematics* (Vol. 1001, pp. 48109-2122). <https://doi.org/10.1145/378886.380416>
24. Shaheen, F., Verma, B., & Asafuddoula, M. (2016). Impact of automatic feature extraction in deep learning architecture. In *2016 International conference on digital image computing: techniques and applications (DICTA)* (pp. 1-8). IEEE.
25. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1), 1929-1958.
26. 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.
27. Vani, S., & Rao, T. M. (2019). An experimental approach towards the performance assessment of various optimizers on convolutional neural network. In *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)* (pp. 331-336). IEEE. <https://doi.org/10.1109/ICOEI.2019.8862686>
28. 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.2017.8119458>