An Efficient Image Classification using SVM and CNN

Sumit Pandey
ABES Enginnering College
Ghaziabad, India

Priya Mishra
ABES Enginnering College
Ghaziabad, India

ABSTRACT

At first glance, the prospect of instructing a computer to perform tasks such as image classification seems highly captivating to our team. Furthermore, there exist several practical applications of this notion in various real-world scenarios. Based on these rationales, our research focus has been directed towards the field of Image Classification. Fortunately, this subject has been extensively investigated by the scientific community, and encountered little difficulty in locating resources for study. Consequently,thoroughly examined a multitude of scholarly articles pertaining to the subject of image classification, each offering a distinct viewpoint. Subsequently, the study made the determination to execute picture classification on a reduced scale due to the constraints imposed by the restricted technology at disposal. Despite the inherent challenges, the study initiated our analysis by employing Support Vector Machines (SVM) in conjunction with a rather limited dataset, ultimately attaining a commendable accuracy rate of 93%. While Support Vector Machines (SVM) is widely recognised as a robust technology, attaining exceptionally high accuracy remains an exceptional occurrence. It was observed that the great accuracy of our results can be attributed to the insufficiency of a sufficiently large dataset. By employing data augmentation techniques, the size of our dataset was expanded by more than threefold. When doing Support Vector Machine (SVM) analysis a notable reduction in accuracy is noted, with a recorded value of 81%. Dissatisfied with the outcomes, the recent works need to explore alternative deep learning methodologies. This inquiry directed our attention towards the study of Neural Networks, namely Convolutional Neural Networks (CNN). By effectively applying Convolutional Neural Networks (CNN), an impressive accuracy of 94.57% was attained on the identical dataset. This serves as evidence of the enhanced capabilities of deep learning approaches compared to conventional machine learning techniques.

Keywords

Python, Support Vector Machines (SVM) and Convolutional NeuralNetworks (CNN)

1. INTRODUCTION

The development of computers or computer programmes that can replicate human intelligence, including perception, reasoning, learning, problem-solving, and decision-making, is referred to as artificial intelligence (AI). Artificial intelligence (AI) is the application of a number of different technologies, including as machine learning, fuzzy[11,13,14,], natural language processing, computer vision, and robotics, to enable robots to carry out tasks that would normally need the intelligence of a human. AI has the ability to revolutionize different industries and enhance efficiency, productivity, and accuracy. Its applications vary from healthcare [10,16] to banking , agriculture[15,17] , education[12] and more, and it has a wide spectrum of applications. Nevertheless, artificial intelligence also brings ethical and societal problems, such as the potential for misuse, employment displacement, bias and discrimination, privacy and security, and potential breaches in these areas. It is of the utmost importance to guarantee that artificial intelligence is created and used properly in order to strike a balance between its potential benefits and threats.

In order to acquire a comprehensive comprehension of picture categorization starting from its fundamental principles. It is advisable to commence with conventional machine learning approaches before transitioning to deep learning techniques.

The objective of this study is to investigate the fundamental concepts and methodologies of picture classification, with a specific focus on elucidating the potential of deep learning techniques and emphasising their significance in comparison to conventional machine learning approaches. As previously indicated, study started by exploring conventional categorization algorithms. After conducting a thorough evaluation of several methods, the work have made the decision to implement Support Vector Machines (SVM). While the utilisation of Neural Networks (NN) appeared to be a more logical choice, given its contemporary nature, the work deemed Support Vector Machines (SVM) to be a more suitable initial approach, and our decision proved to be accurate. Prior to discussing the concept of Support Vector Machines (SVM) and our implementation approach, it is necessary to provide an explanation for our initial consideration of Neural Networks (NN) as a potential alternative. The accuracy of image categorization achieved by neural networks surpasses that of any other machine learning technique, including other deep learning algorithms. Traditional machine learning algorithms exhibit a point of saturation in their performance, whereby the inclusion of additional data beyond this threshold does not yield any further improvements in performance. Having discussed the significance of deep learning algorithms, the study may now proceed to examine support vector machines (SVM). Upon initial comprehension of this methodology, It should be noted that Support Vector Machines (SVM) is a highly robust classification technique based on its inherent strengths. However, as previously indicated, it lacks viability in a world abundant with data. Upon conducting additional investigation.

The implementation of a three-fold categorization of the subject matter facilitated a more effective comprehension. The focal areas encompassed within the broader topic are those related to visualization. The Optimization of Parameters and the Execution and Evaluation of the Process. When provided with a training dataset that has been labelled, a supervised learning algorithm produces an ideal hyperplane. This hyperplane represents a line that has undergone transformation through a process known as kernelling. Despite the development of a classifier, its efficiency remains suboptimal. The Support Vector Machine (SVM) algorithm employs the greatest gap trick optimisation technique to identify the optimal classifier. In this process, support vectors play a crucial role in the computation of the optimal classifier. Therefore, more points do not contribute to the argument. Following the aforementioned procedures, the achieved accuracy did not meet our initial expectations. Consequently, the work conducted further investigation and arrived at the idea of employing Deep Learning, a technique that enhances learning by modifying the synaptic weights to effectively transmit input signals across multiple layers to neurons linked with relevant overarching concepts. When input data is introduced into a neural network, each activated artificial neuron propagates signals to specific neurons in the subsequent layer. The potential firing of neurons in response to the reception of numerous impulses is a likely occurrence. The complete procedure effectively eliminates extraneous signals and preserves solely the most pertinent characteristics. Furthermore, the attributes provided to a neural network model consist just of the pixel values associated with the image. The utilization of this technique establishes a level of abstraction that accommodates those with limited expertise in the field of image processing. Significantly, the autonomous capacity to acquire traits positions it in a distinct category. The subsequent sections of the paper are structured in the following manner. Section 2 of the document presents an overview of the relevant literature and previous research conducted in the field. In this paper, Section 3 introduces the proposed Solution, while Section 4 provides a detailed explanation of the Performance Analysis. The findings are presented in Section 5. The section 6 present the results which is followed by conclusion and outline potential avenues for future research.

2. RELATED WORK

The purpose of this article [1] is to investigate and provide a condensed grasp of a variety of picture classification strategies and different classification techniques. In addition to this, the paper investigates a variety of methodologies that are dependent on various kinds of inputs. The author argues, in the second part of the article, that the process of classifying images, which serves as the last stage of the pattern-matching procedure, is an important aspect of pattern recognition. In addition, the technique for classifying data serves the purpose of outlining the degree of precision involved in the recognition of patterns. In reference [3], the author discusses the significance and complexity of classification within the field of computer science. Classification is determined by various factors, including the description and likeness of items. In summary, the pixel serves as the fundamental unit for classification and is organised into many classes. In reference [4], the Convolutional Neural Network (CNN) is regarded as the most advanced technology for the categorization of traffic signs. Furthermore, the outcomes obtained from the use of Convolutional Neural Networks (CNN) serve the purpose of acquiring knowledge about distinctive attributes and categorising RGB-D photographs. The transfer learning technique known as "Fine-tuning Technique" is employed, wherein pre-trained layers are reused to categorise the testing set. In reference [5], the author asserts that in the current era, there is a notable surge in the popularity of picture classification across multiple domains, including image processing, computer vision, and machine learning. This research primarily centres on the utilisation of deep learning techniques for the purpose of addressing the image classification challenge. In reference [6], a dataset consisting of 1.2 million high resolution photos is subjected to classification by a Convolutional Neural Network (CNN) that has been trained. The CNN successfully categorises the images into 1000 distinct classes, as part of the ImageNet benchmark. The neural network is comprised of five convolutional layers, each containing max-pooling layers, as well as three fully connected layers. The total number of neurons in this network is 650,000. In reference [7], artificial neural networks (ANN) and support vector machines (SVM) are two often utilised techniques in the field of image categorization. The responsive class exhibited by an artificial neural network (ANN) is the outcome of the classification process applied to a sub-image. The experimental findings demonstrate that the recognition application achieved a precision rate of 86%. This model is referred to as ANN SVM, as it incorporates multiple artificial neural networks (ANNs) and one support vector machine (SVM) within the proposed framework. According to reference [8]. The preceding scholarly articles examined the fundamental methodologies employed in the field of picture categorization. The examination of these strategies provided us with sufficient understanding to progress towards more sophisticated methodologies. The survey study under consideration classifies image classification techniques into two main categories: supervised learning and unsupervised learning. Specifically, the techniques discussed in this research encompass Artificial Neural Networks and Support Vector Machines. The Directed Acyclic Graph (DAG) is a visual representation employed to depict the interconnections among variables or nodes inside a given system. Within the domain of machine learning, the directed acyclic graph (DAG) can be employed as a tool for the purpose of representing and examining a wide range of algorithms and methodologies. One technique that can be employed is the Support Vector Machine (SVM), which is classified as a form of supervised learning. The Support Vector Machine (SVM) technique is widely employed in both classification and regression tasks because to its robust capabilities. The objective is to identify an ideal hyperplane that effectively separates distinct classes within the feature space. In contrast, Fuzzy Decision Trees exemplify unsupervised learning

methodologies. In contrast to Support Vector Machines (SVM), which necessitate the availability of labelled data for training purposes, Fuzzy Decision Trees operate independently of predetermined class labels. In contrast, fuzzy logic is employed. This talk will analyse three classification systems, specifically. Supervised learning is the acquisition of knowledge regarding the behaviour of a function through the examination of several input-output examples of such function. Within the domain of machine learning, the concept of unsupervised learning pertains to a framework wherein the learning algorithm does not receive labelled instances as guidance for acquiring a function. The lack of annotated data is a difficulty for the learning procedure, as it requires the algorithm to independently recognise patterns and structures within the data in order to extract valuable insights. Semi-supervised learning pertains to a machine learning methodology wherein the training dataset encompasses an amalgamation of annotated and unannotated data.

The study commenced by gathering a dataset comprising 350 photographs, all of which were in a pristine condition with minimal noise. The initial implementation, which employed Support Vector Machines (SVM) for image classification, demonstrated satisfactory outcomes due to the utilisation of this machine learning technique. The dataset employed in this study was deemed sufficient for achieving favourable findings. A level of accuracy of 93% was achieved. However, similar findings were not detected in cases where the photos exhibited slight blurring or had a higher level of brightness in the backdrop. It was determined that the CNN (Convolutional Neural Network) would be employed for the aforementioned purpose. One significant challenge associated with utilising a Convolutional Neural Network (CNN) is the substantial data requirement for effectively training the model. In this study, various data augmentation approaches were examined in order to enhance the dataset by increasing the quantity of photos. Several other photos were included. Additionally. A cumulative quantity of 3000 pictures was produced. The employed data augmentation techniques encompassed the conversion of the colour space from RGB to CMY, as well as the application of image dithering. Generating a rendition of the image in primary colours. The image is subject to a random translation. Inverting the image along the vertical axis. Performing a horizontal image flip. The initial step in data analysis is preprocessing. This involves cleaning and transforming the raw data to the commencement of the project, a dataset including 100 photographs was gathered. Furthermore, all of the photos remained undamaged and had minimal levels of noise. Each image obtained from the dataset undergoes the various processes described in the preceding section, resulting in the creation of more than 7-8 images that encompass all the aforementioned activities. The collection comprises a total of over 800 photos. The dataset is subsequently imported into Python using a custom-defined function. This method initially resizes all the images to dimensions of 64 by 64. Subsequently, the images are flattened using the appropriate algorithms from the NumPy library, resulting in an array containing numerical representations of each image. The aforementioned photos or the collection of arrays are then utilised for the purposes of processing and categorization. This form of data is highly valuable due to its conversion of digital images into numerical representations, enabling efficient processing for tasks such as image classification utilising techniques like Support Vector Machines (SVM), Convolutional Neural Networks (CNN), or other applicable methods.

3. DECISION AND CLASSIFICATION

The study initially focused on traditional machine learning techniques for classification, with SVM being chosen for implementation. Although Neural Networks (NN) seemed like a more modern and logical choice, SVM was selected as a starting point, and this decision proved to be effective. Before discussing what SVM is and how it was implemented, it is important to explain why NN appeared to be a strong candidate. The accuracy of NN in usage classification surpasses other learning algorithms, including deep learning approaches. Additionally, NN's ability to autonomously learn features sets it apart. As illustrated, traditional machine learning algorithms eventually reach a performance plateau, where adding more data doesn't improve results. This is where NN excels. Moreover, with NN, the input features can simply be pixel values, making it accessible even for those with limited knowledge of image processing. The significance of deep learning algorithms is thus emphasized.

4. PERFORMANCE ANALYSIS

Once the work comprehended all the tweaking parameters,the need to prepare execute the Support Vector Machine (SVM) algorithm. There are two distinct implementations, each utilising a unique dataset. The initial implementation was utilising the IRIS Dataset as a means to comprehend the practical use of Support Vector Machines (SVM). After gaining comprehension of the methodology, the work proceeded to apply the Support Vector Machine (SVM) algorithm on the dataset previously indicated, specifically employing SVM with Image Pixel Values as Features. It should be noted that during the implementation of Support Vector Machines (SVM), our dataset consisted of RGB images belonging to five distinct classes. Every image is kept in the format of '[class name] [serial no].jpg'. The dataset consists of 67 photos featuring dalmatian dogs at various stages of development, encompassing both animated and non-animated representations, and including variations with and without background elements. The photos of 53 dollar bills are resized to match the dimensions of the bill. A total of 53 pizza slice photos were obtained by cropping them from commercial posters. With and without context. There are a total of 65 photos depicting soccer balls. A total of 85 photographs of sunflowers were captured in their natural environment, showcasing both the full plant and specific details. After applying the aforementioned procedures, the dataset consists of a total of 323 images. The classes within the dataset are marked numerically, ranging from 0 to 4. One may inquire as to the rationale behind selecting Support Vector Machines (SVM) as opposed to employing basic clustering techniques. Principal component analysis (PCA) was utilised to reduce the dimensionality .

.

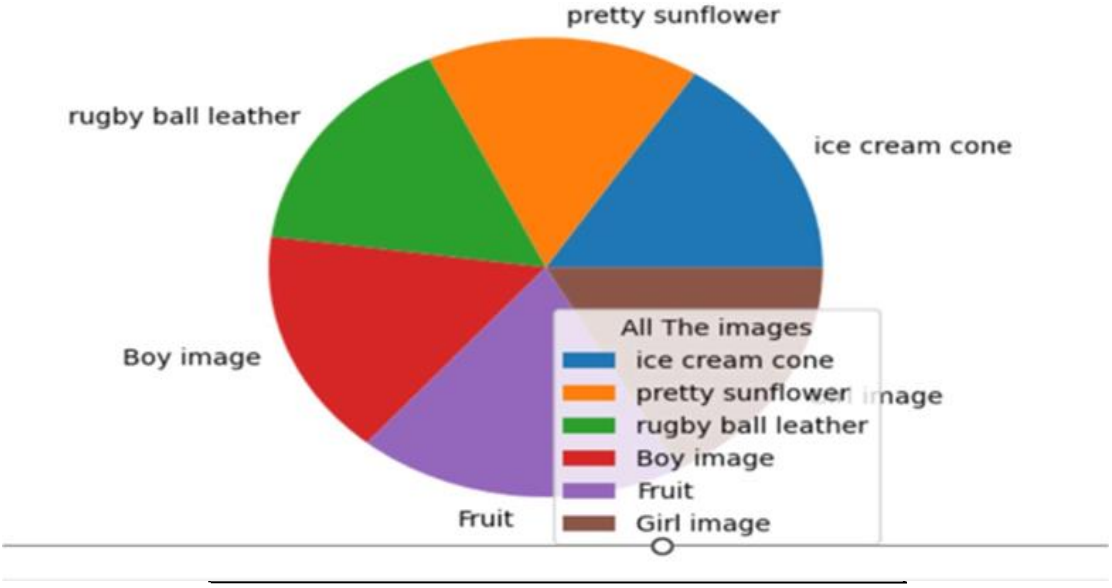

Fig 1:Graph illustrating Disadvantages of clustering

During the implementation phase, two kernels were utilised. Initially, the linear kernel was employed due to the lack of linear separability in the data. However, the accuracy achieved with this kernel was notably poor. Consequently, a transition was made to the radial basis function (RBF) kernel, which unsurprisingly yielded a higher fidelity index (FI) score. The trials were conducted using C values of 11, 10, 100, and 1000, as well as gamma values of 0.001 and 0.0001.

5. METHODOLOGY:

The classification of images using Convolutional Neural Networks (CNN) involved the implementation of the following steps. Following the successful implementation of Support Vector Machines (SVM), our focus shifted towards Neural Networks, specifically Convolutional Neural Networks (CNN). The architecture of the CNN utilised in this study is outlined below:

The methodology employed in this study involves the utilisation of the LeNet5 architecture, which shares similarities with the structure of Convolutional Neural Networks (CNNs).

1. Initially, a picture with dimensions of 120x120x1 is subjected to convolution with a filter, resulting in an output of size 116x116x6.
2. The third component is the output. The application of a max-pooling operation with a stride and size of 2 reduces the output from 4 to a dimension of 58x58x6.
3. The output is further reduced to a size of 27x27x16 with the application of a filter and max-pooling with a stride and size of 2.The outcome is subsequently compressed until a completely linked layer of 84 neurons is achieved. Finally, the softmax function is utilised to obtain the output probabilities for the five image classes, indicating the likelihood of each class membership.

6. RESULTS

The different data augmentation techniques used are;

Conversion to CMY Colour space: The images in the dataset are in RGB color space and all the 350+ images are converted to CMY colour space. It was done using the user defined function which is shown below.

Fig.2: Conversion to CMY Colour space:

6.1 Dithering of the Image to add noise to the image

In the field of image processing, dithering refers to a technique that simulates a wider variety of colours and tones than a given device is capable of displaying on its own. It is a technique that involves adding noise to an image in order to give the impression that there are more colours present than there actually are. Applications that require digital image processing, such as printing, display, and photography, frequently make use of the dithering technique which is followed by primary color creation, translation and flipping. The results are displayed in figure()

(a)Dithering of the image

(b)Primary Colors Version the image

(c) **Translation of the image:** **(d) Flipping the image**

image

The figure 4 shows the results in terms of loss and accuracy of models. with respect to epochs

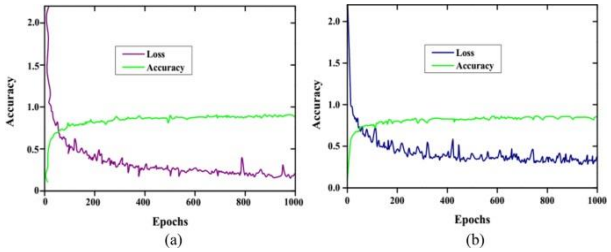

Fig.5: Accuracy and losses of models with epoch

Table 1. Accuracy Assessment of CN

Item	Dalmatian	Dollar-Ball	Pizza	Sunflower
Precision	0.94	1	0.83	0.91
Recall	0.86	0.94	0.98	0.92
F1-Score	0.89	0.96	0.93	0.93

7. CONCLUSION AND FUTURE WORKS

Even though SVM is a very powerful technique, therefore, the study aims to implement it with a relatively minimal dataset and still get an accuracy of 93%. Attaining such a significant level of precision remains an exceptional occurrence. Ultimately, it was discerned that the elevated precision of our findings can be attributed to the insufficiency of a sufficiently extensive dataset. By employing data augmentation techniques, the size of our dataset was expanded by more than threefold. Subsequently, upon reapplying the Support Vector Machine (SVM) algorithm,a notable decline in accuracy, with a reported value of 81% is recoreded. Dissatisfied with the outcomes, the need to explore alternative deep learning methodologies. This inquiry directed our attention towards the study of Neural Networks and Convolutional Neural Networks (CNNs). Upon effective implementation of Convolutional Neural Networks (CNN), an impressive accuracy of 94.57% was attained on the identical dataset. This serves as evidence of the enhanced capabilities of deep learning approaches compared to conventional machine learning techniques. Convolutional Neural Networks (CNNs) share similarities with conventional neural networks as they comprise hidden layers composed of neurons with adjustable parameters. Typically, CNNs employ the SoftMax function as the classifier in the final layer of the network. Nevertheless, it is important to acknowledge that our research has identified other papers that present contrasting findings. The present research present the application of Support Vector Machine (SVM) within an artificial neural network framework. Given our current understanding and proficiency in implementing both Convolutional Neural Networks (CNN) and Support Vector Machines (SVM), the work evaluates and contributes to existing research in these areas.

8. REFERENCES

[1] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105). Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. Link

[2] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) (pp. 770-778). Link

[3] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) (pp. 248-255). Link

[4] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Fei-Fei, L. (2015). ImageNet large scale visual recognition challenge. International Journal of Computer Vision, 115(3), 211-252. Link

[5] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) (pp. 1-9). Link

[6] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. International Journal of Computer Vision, 115(3), 211-252. Link

[7] Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572. Link

[8] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016)

[9] Mishra, Priya, et al. "Early Predication of Covid-19 by Machine Learning Algorithms." *Journal of Pharmaceutical Negative Results* (2022): 2907-2914.

[10] Mishra, Priya, Charu Gandhi, and Buddha Singh. "Link quality and energy aware geographical routing in MANETs using fuzzy logics." *Journal of Telecommunications and Information Technology* (2016).

[11] Verma, Shikha, et al. "STUDENT's BEHAVIOR AND REGULARITY EFFECTS RESULT."

[12] Mishra, Priya, Saurabh K. Raina, and Buddha Singh. "Effective fuzzy-based location-aware routing with adjusting transmission range in MANET." *International Journal of Systems, Control and Communications* 7.4

(2016): 360-379.

[13] Mishra, Priya, Charu Gandhi, and Buddha Singh. "An improved greedy forwarding scheme in MANETs." *Journal of Telecommunications and Information Technology* (2017).

[14] Mishra, P., Raina, S. K., & Singh, B. (2016). SDLAR: self adaptive density-based location-aware routing in mobile adhoc network. In Advanced Computing and Communication Technologies: Proceedings of the 9th ICACCT, 2015 (pp. 401-409). Springer Singapore.

[15] Gupta, M., Jain, S. (2023). Banana Leaf Diseases and Machine Learning Algorithms Applied to Detect Diseases: A Study. In: Dutta, P., Chakrabarti, S., Bhattacharya, A., Dutta, S., Shahnaz, C. (eds) Emerging Technologies in Data Mining and Information Security. Lecture Notes in Networks and Systems, vol 490. Springer, Singapore. https://doi.org/10.1007/978-981-19-4052-1_14.

[16] P. Mishra and S. K. Prasad, "Reviewing Flood Susceptibility Mapping Utilizing Machine Learning: Insights and Prospects for Future Research," 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2024, pp. 649-655, doi: 10.1109/ICAAIC60222.2024.10575530.