A Survey and Comparison of Artificial Intelligence Techniques for Image Classification and Their Applications

Vishal .T .V[1], Srinidhi .S[2], Srividhya .S[3], Sri Vishnu Kumar .K[4], Swathika R[5]

[1]Department of Computer Science and Engineering, SSN College of Engineering, Kanchipuram, Tamil Nadu, India

[2, 3, 4, 5]Department of Information Technology, SSN College of Engineering, Kanchipuram, Tamil Nadu, India

Abstract: *This paper focuses on a case study of a number of image classification algorithms which include decision trees, k-nearest neighbours, deep neural networks, Convolutional neural networks, Support vector machines and random forest. The three main algorithms – SVM, RF and CNN have been elaborated due to their varying approaches to classification. They have been juxtaposed with their applications in various domains which are an influence in technology. This paper also discusses the multiple advantages and disadvantages of the aforementioned algorithms.*

Keywords: Artificial Intelligence; Image Classification; Convolutional Neural Networks; Support Vector Machines; Random Forests*;*

1. Introduction

In the present world, the interaction between humans and computing devices has advanced to an extent where it has become a necessity. The technology is embedded with our daily schedules as it helps us work, communicate, lay out business plans with ease. The popular notion is that the existing techniques in the domain of computing, communication and display technologies sometimes serve as a bottleneck in effective utilization of the available information flow. To efficiently use them, these applications require a great deal of interaction with the user. This is where the concept of artificial intelligence comes into play. Artificial intelligence is a humongous field wherein the computers are trained to exhibit intelligent behavior. This effectively reduces the human labor and assists them in creating several paradigms. One major factor to consider when creating an intelligent system is to perceive how they deal with the changing environments and navigate through them successfully. Hence, computer vision provides this kind of information in an optimized manner and serves as a vision sensor.

Computer vision is a process of acquiring, processing and analyzing images in the real world to display some numeric information about them. Therefore, object recognition is the task of recognizing the three-dimensional or, in general, multi-dimensional objects from two-dimensional images and classifying them into one of the many known object types. Humans are said to perform object detection and recognition effortlessly and instantaneously. A person can identify a multitude of objects each with different viewpoints, colors, textures and sizes with minimal efforts provided the objects are visible to the person. This is the exact idea which is currently being used in computer vision. The objects are recognized by the intelligent systems based on the various features that they possess and further are classified to produce great efficiency.

Object recognition is closely related to many computer applications such as robotics, surveillance etc. A number of object recognition algorithms and intelligent systems have been proposed to make this process a more optimized one. This paper discussed the various methods which provide consistent and precise recognition of objects in images.

2. Literature Survey

Classification algorithms such as decision trees, random forests, and SVM are statistical methods for grouping data. These data may be words, colors, sounds etc. In an image classification context, these algorithms can be used as pixel-based classifiers. The features given as input to the classifiers is the raw pixel values of the images. However, the image size must be the same for all the training and testing images to maintain a constant number of features.

2.1 Support Vector Machines

A Support Vector Machine builds ahyperplane or a set of hyperplanes in high or infinite-dimensional space, used for classification. It is a maximum margin classifier as the algorithm generates decision boundaries that have the largest distance to the nearest training data feature vector [1]. The feature vectors that are closest to the decision boundary are called support patterns of the decision boundary. To keep the computational load reasonable, the mappings used by SVM schemes are designed to ensure that the dot products may be computed easily in terms of the variables in the original space, by defining them in terms of a kernel function k(x, y) selected to suit the problem [2]. Hyperplanes in the higher-dimensional space are defined as the set of points whose dot product with a vector in that space is a constant [3]. SVMs show lesser generalization error as it is a maximum margin classifier. The performance and accuracy depend on the hyperplane selection, number of features and the type kernel used.

2.2. Decision Tree

A decision tree is a hierarchical classifier that uses a set of rules to segment the predictor space. The learning process

involves creating the set of rules for the dataset [4]. A decision tree has two types of nodes, leaf nodes and internal nodes. Leaf nodes have a class label determined by a majority of the training samples reaching that leaf and internal nodes represent a test on an attribute depending on which child nodes branch out [4]. One of the best features of decision trees is to break down complex decision-makingprocess into a set of simpler decisions. Construction of Decision tree classifiers can be divided into four categories: Bottom-up approaches, Top-down approaches, hybrid approaches and Growing Pruning approaches [5]. The most commonly used approach is the top down approach which involves selection of a node splitting rule, the decision as to which node is terminal and a class label for the terminal node [5]. Entropy is a measure of the unpredictability of information content and at each stage of the tree, the focus is to maximize entropy reduction [5].To avoid overfitting of a decision tree Pruning is done ,which basically reduces the size of the tree by eliminating irrelevant nodes[6]. Shepherd B.A. [7] used ACLS (Analogue Concept Learning System) algorithm to generate a decision tree which uses both numerical and logical attributes to classify 12 different types of chocolate images and proved that its performance is better than K-nearest neighbor classifier and Maximum distance classifier. The ACLS trees were found to be advantageous in terms of cost of making a classification and intelligibility of solution [7].

2.3 Random Forests

A Random Forest Classifier consists of a collection of decision tree classifiers where the decision trees are independently distributed random vectors and each tree casts a unit vote for the most popular class for a given input [8]. The general method of random decision forests was first proposed by Ho in 1995, who established that forests of trees splitting with oblique hyperplanes if randomly restricted to be sensitive to only selected feature dimensions, can gain accuracy as they grow without suffering from overtraining [9]. Random forest classifiers use a modified tree learning algorithm in which the concept of bagging is used in tandem with random feature selection [9]. Typically, for a classification problem with p features, features are used in each split. This method displays better performance models than regular decision trees as the variance is decreased without increasing the bias. The average of many trees is not sensitive to noise as opposed to a single tree which is highly sensitive to noise. The explanation of the forest method's resistance to overtraining can be found in Kleinberg's theory of stochastic discrimination [10] [11] [12].

2.4 K-Nearest Neighbours

K-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification [13]. When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant then the input data will be transformed into a reduced representation set of features (also named features vector) [13]. It is a non-parametric model as stated by N.S. Altman in his work on KNNs [14]. In this model, the new data point, is compared with k nearest sample datapoints, and the class with a maximum number of nearest neighbours to the new data point is deemed as the class of the datapoint. It is generally used for text-based analysis and sometimes fast image classification after the features of the image are extracted using Principle Component Analysis.

2.5 Deep Neural Networks

A deep neural network (DNN) is an artificial neural network (ANN) with multiple hidden layers of units between the input and output layer [15]. Searching the parameter space of deep architectures is a difficult task, but learning algorithms such as those for Deep Belief Networks have recently been proposed to tackle this problem with notable success, beating the state-of-the-art in certain areas[16].Hinton et al. recently proposed a greedy layer-wise unsupervised learning procedure relying on the training algorithm of restricted Boltzmann machines (RBM) to initialize the parameters of a deep belief network (DBN), a generative model with many layers of hidden causal variables. This was followed by the proposal of another greedy layer-wise procedure, relying on the usage of auto-associator networks. [17] The predominant methodology in training deep learning advocates the use of stochastic gradient descent methods (SGDs).More sophisticated off-the-shelf optimization methods such as Limited memory BFGS (L-BFGS) and Conjugate gradient (CG) with line search can significantly simplify and speed up the process of pretraining deep algorithms. [18]

2.6 Convolutional Neural networks

Convolutional neural networks are best suited for image recognition, as they retain the spatial topology [19]. The general strategy of a convolutional network is to extract simple features at a higher resolution, and then convert them into more complex features at a coarser resolution [20]. They use three main architectural ideas to ensure some degree of scale, shift and distortion: Local receptive fields, shared weights and spatial or temporal sub-sampling [19].Though Convolutional neural networks are best suited for image recognition, they are not used in the engineering community, due to their complex implementation [19]. A recent achievement in deep learning is from the use of convolutional deep belief networks (CDBN). A CDBN is very similar to normal Convolutional neural network in terms of its structure. Therefore, like CNNs, they are also able to exploit the 2D structure of images combined with the advantage gained by pre-training in Deep belief network. They provide a generic structure which can be used in many image and signal processing tasks and can be trained in a way similar to that for Deep Belief Networks. Recently, many benchmark results on standard image datasets like CIFAR have been obtained using CDBNs [21].

3. Algorithm Explanation

Figure 1: Classification Block Diagram

3.1 Convolutional Neural Networks

Convolutional neural networks (CNNs) are feed forward artificial neural networks. They mimic the multilayer perceptron model of the human eye, by using multiple filters at each layer. They are designed to use minimal preprocessing. CNNs use multiple layers of small neuron collections, which take multiple small portions of the image and this helps them to deal with image translations. CNNs have spatially arranged neurons, which are 3 dimensional in nature, namely, height, width and depth. The spatial effect comes in when these neurons are connected only to a few neurons in a previous layer instead of the entire set of neurons in the layer before it.

CNN exhibits a concept called shared weights, according to this concept, each filter h_i is applied across the entire visual field. This results in shared parameterization and results in the creation of a feature map. This replication helps in detecting features in an image, irrespective of their location. A feature map is obtained by convolving a filter with spatially related pixels. Convolution is performed by multiplying the pixels with a linear weight function, and then adding a bias, followed by conversion to a non-linear function.

$$h_{ij}^k = tanh((W^k * x)_{ij} + b_k$$

Putting it all together, W_{ij}^{kl} denotes the weight connecting each pixel of the k-th feature map at layer m, with the pixel at coordinates (i,j) of the l-th feature map of layer (m-1).An important concept of down sampling is pooling, which is non-linear down sampling. The intuition is that, only the maximum value of the pooled set of pixels is taken, because, only the spatial location is important, and not the exact location of the image. This results in decreasing the computational complexity and also the number of features. Feature map size decreases with depth, layers near the input layer will tend to have fewer filters while layers higher up can have more. To equalize computation at each layer, the product of the number of features and the number of pixel positions is typically picked to be roughly constant across layers. The various applications of CNNs are image recognition, video analysis and natural language processing.

Figure 2: Convolutional Neural Networks Model

3.2 Support Vector Machines

Support vector machine is a non-probabilistic binary classifier which when given some training data outputs an optimal hyperplane which classifies new samples. Samples can be sometimes separated into two classes just by drawing a line in-between and such type of data are called linearly separable data and the line used to separate the two classes is called a decision boundary. But if the original problem stated in a finite dimensional space cannot be separated in that space, it can be mapped into a high or infinite dimensional space where it can be easily classified using a hyperplane as the decision boundary. Training samples that support the decision boundary are called support vectors. To avoid explicit mapping to higher dimensions we can use special types of functions called kernels which reduce the computational complexity of the algorithm.

If we have a data point viewed as a p-dimensional vector separated by a p-1 dimensional hyperplane the classifier is known as linear SVM classifier. In SVMs, Margin is the band around the decision boundary without any training samples and maximizing the margin is important because it will reduce the test error and avoid over-fitting. Any hyperplane can be represented using a set of points $\vec{x}$ as $\vec{w}.\vec{x_i} - b = 0$ where $\vec{w}$ is a normal to the hyperplane. If the data is linearly separable, we can draw two parallel hyperplanes on either side of the decision boundary separating the two classes. We can maximize the distance between them which is the margin. The equations of the two parallel hyperplanes are
$\vec{w}.\vec{x_i} - b = 1$ And $\vec{w}.\vec{x_i} - b = -1$.
To prevent the points from going to the wrong side of the boundary we have the following constraint,

$y_i(\vec{w}.\vec{x_i} - b) \geq 1$ For all $1 \leq i \leq n$

Where y_i gives the class of the data. The geometrical distance between the hyperplanes is given by $\frac{2}{\|\vec{w}\|}$. So to maximize the margin we need to minimize $\|\vec{w}\|$ subject to

$$y_i(\vec{w}.\vec{x_i} - b) \geq 1 \text{ for i} = i \dots n \qquad (2)$$

By using lagrangian multipliers we get the simplified dual form of the problem as:

$$Maximize \; f(c_1 \dots c_n) = \sum_{i=1}^{n} c_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i c_i (x_i.x_j) y_j c_j$$

$Subject\ to\ \sum_{i=1}^{n} c_i y_i = 0\ and\ 0 \leq c_i \leq \frac{1}{2n\lambda}\ for\ all\ i$

The equation is given to a QP solver which gives out c_i (lagrangian multiplier) which when non-zero corresponds to a support vector. The optimal hyperplane is found using the equation

$$\vec{w} = \sum_{i=1}^{n} c_i y_i \vec{x_i} \qquad (4)$$

For non-linear classification, we have to use a mapping function to map to a higher dimension, where a maximum margin hyperplane can be fit and data can be classified. We can use the kernel trick here such that $k(\vec{x_i}, \vec{x_j}) = \varphi(\vec{x_i}).\varphi(\vec{x_j})$, Where $\varphi(\vec{x_i})$ is the mapping function. Generalization error increases while working in higher dimensional space but when given enough number samples the algorithm works fine. So the dual form mentioned above will be slightly different involving the kernel function, instead of dot products.

$$Maximize\ f(c_1 \dots c_n) = \sum_{i=1}^{n} c_i + \frac{1}{2}\sum_{j=1}^{n} y_i c_i \left(\varphi(\vec{x_i}).\varphi(\vec{x_j})\right) y_j c_j$$

$$= \sum_{i=1}^{n} c_i + \frac{1}{2}\sum_{j=1}^{n} y_i c_i k(\vec{x_i}.\vec{x_j})\, y_j c_j$$

$Subject\ to, \sum_{i=1}^{n} c_i y_i = 0\ and\ 0 \leq c_i \leq \frac{1}{2n\lambda}\ for\ all\ i$ (4)
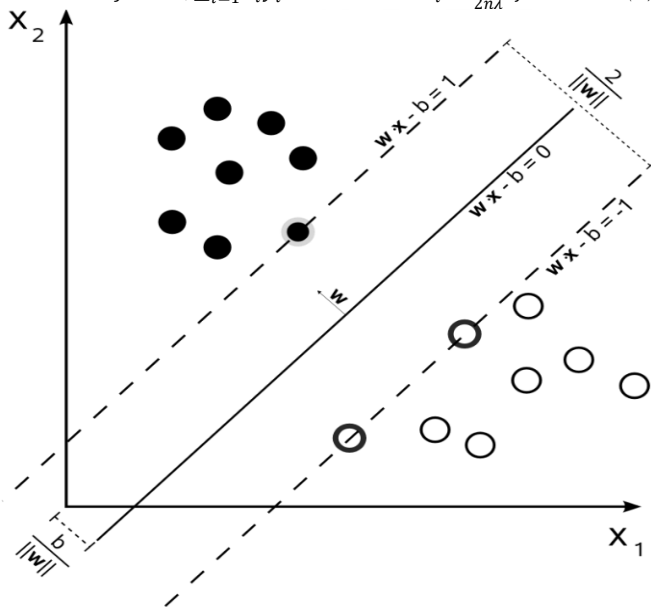
Figure 3: Maximum Margin Hyperplane separating two classes [3]

Polynomial kernel and Radial Basis Kernels are some of the popular kernels.

Polynomial kernel,
$$k(\vec{x_i}, \vec{x_j}) = (\vec{x_i}.\vec{x_j} + 1)^d$$

Gaussian radial basis kernel,
$$k(\vec{x_i}, \vec{x_j}) = exp\left(-\gamma \|\vec{x_i} - \vec{x_j}\|^2\right)\ for\ \gamma > 0$$

(3)

SVMs are used in pattern recognition, object detection and bioinformatics. SVMs are extensively used for recognizing handwritten characters. For image classification, we can choose different types of features like raw pixels or histogram of oriented gradients (HOG) to train the classifier. The image is divided into cells and for the pixels inside the cell a histogram of gradient direction is computed, thus giving the count of gradient orientation across localized portions of the image. Concatenation of the histograms gives the descriptor. This method is especially suited for human detection in images and videos. SVMs require full labelling of a data and they can directly be applied only for two class tasks.

3.3 Random Forests

A random forest classifier is an ensemble learning method for classification, regression and other tasks that operate by constructing a number of decision trees at training time and outputting the class that is the mode of the classes of the individual trees. The mean prediction is used in the case of regression. It was shown that the forest error rate depends on two things [1]:

- The correlation between any two trees in the forest, increasing the correlation increased the forest error rate.
- The strength of each individual trees in the forest. A tree with a low error rate is a strong classifier. Increasing the strength decreases the forest error rate.

When the training set for the current tree is drawn by sampling with replacement, about one-third of the cases are left out of the sample. This out-of-bag data is used to get a running unbiased estimate of the classification error as trees are added to the forest. It is also used to get estimates of variable importance. After each tree is built, all of the data are run down the tree, and proximities are computed for each pair of cases. If two cases occupy the same terminal node, their proximity is increased by one. At the end of the run, the proximities are normalized by dividing by the number of trees. Proximities are used in replacing missing data, locating outliers, and producing illuminating low-dimensional views of the data. In random forests, there is no need for cross-validation or a separate test set to get an unbiased estimate of the test set error. It is estimated internally, during the run, as follows:

- Each tree is constructed using a different bootstrap sample from the original data. About one-third of the cases are left out of the bootstrap sample and not used in the construction of the kth tree.
- Put each case left out in the construction of the kth tree down the kth tree to get a classification. In this way, a test set classification is obtained for each case in about one-third of the trees. At the end of the run, take j to be

the class that got most of the votes every time case n was Out Of the Bag (OOB). The proportion of times that j is not equal to the true class of n averaged over all cases is the oob error estimate. This has proven to be unbiased in many tests.

Given a training set $X = x_1, x_n$ with responses $Y = y_1... y_n$, bagging repeatedly (B times) selects a random sample with replacement of the training set and fits trees to these samples:
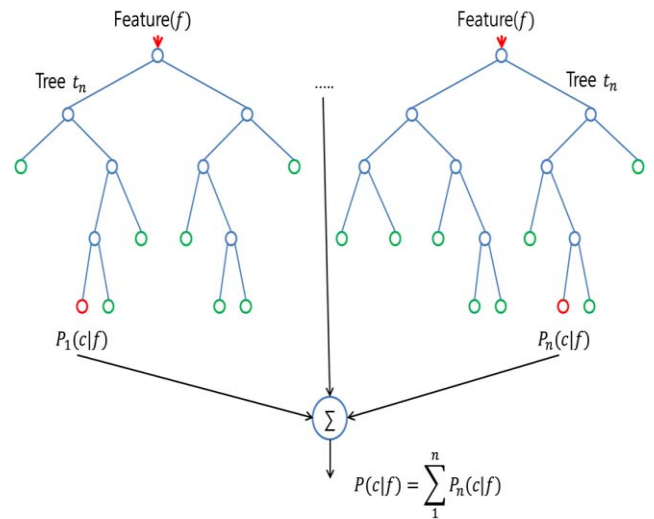
For $b = 1... B$:

1) Sample, with replacement, n training examples from X, Y; call these X_b, Y_b.
2) Train a decision or regression tree f_b on X_b, Y_b.

After training, predictions for unseen samples x' can be made by averaging the predictions from all the individual regression trees on x':

$$\hat{f} = \frac{1}{B}\sum_{b=1}^{B}\hat{f}_b(x')$$

or by taking the majority vote in the case of decision trees.

Figure 4: Random Forest Structure

(7)

Table 1: Comparison of Algorithms

Classification Algorithm	Feature Extraction (Yes/No)	Advantages	Disadvantages	Applicatiions
SVM	No	• Kernel can perform either linear or non-linear transformation • The problem of overfitting is eliminated. • Computational Complexity is greatly reduced. • The choice of form of threshold provides flexibility. • Provides good generalization capability.	• Classifier Training consumes a lot of time. • Multi-Class Classifier tougher to implement compared to Binary Classifier. • Tough to determine optimal parameters in non-linear separation of data. • Result transparency is low.	• SVMs are widely used in human detetion. optical character recognition, medical image analysis and pattern recognition.
CNN	Yes	• Highest accuracy of image classification amongst all algorithms • Designed for minimal preprocessing • Pooling decreases complexity • Relearning is possible	• Nonlinear time complexity • Takes long to train • Because of minimal preprocessing, initial stages take a lot of time • May overfit data sometimes	• Convolutional Neural Networks is the most most popular algorithm for image recognition and video analysis. They are also used in natural lanuage processing and drug discovery which involves predicting interaction between molecules and biologial proteins.
Random Forests	No	• Not that sensitive to overfitting. • Selection of random subset of features for each tree increases classification accuracy. • More robust than a decision tree. • It can compute proximity between samples. • Variable importance for the classification can be computed. • Faster to train.	• Run time performance is affected while working with very large number of features. • Visual interpretation of the model is very difficult. • For data including categorical variables with different number of levels, Random forests are biased towards the attributes with more levels.	• Random forests have been used in face detectin, object detection and classification of reamotely sensed images. • Microsoft has used random decision forests for human pose Recognition in their motion senesing input device ,Kinect.

4. Applications

Artificial intelligence tools are applied to a variety of domains which include medical image classification, remotely sensed image classification, face recognition and object detection. The various classifiers which are used in the aforementioned fields have been elucidated in this section.

4.1 Medical Image Analysis

Machine learning is an important tool in the field of medicine, as; it helps classify medical images accurately. X-ray image classification uses random forests with local wavelet based local binary pattern to improve classification

Volume 5 Issue 4, April 2016

191

performance as well as training and testing time.To classify the X-ray images, local wavelet based Centre Symmetric Local Binary Pattern (WCS-LBP) descriptors from local parts of the image are extracted, which, are then applied to the extracted feature vector to create decision trees and form a random forest. [22]Another such approach to medical image classification is done by high dimensional template warping and morphological representation of the anatomy of interest. Consequently, watershed segmentation is used to extract regions that display thestrong correlation between morphological measurements and classification (clinical). Support Vector Machine- Recursive feature Elimination is used to rank computed features and finally SVM is applied using the best set of features. This method showed 91% accuracy in the classification of 61 brain images of female with normal control and schizophrenia patients. [23] SVM's have also been used to classify Alzheimer's disease from normal aging in individual scans. Up to 96% of the pathologically verified AD patients were correctly classified using a set of brain images. Differentiation between people who are suffering from AD from those who are suffering from front temporal lobar degeneration was also possible. [24]

4.2 Remote Sensed Image Analysis

Remote sensing is another field which uses classification algorithms. In one such attempt, univariate, multivariate and hybrid decision trees were used to classify land cover from remotely sensed data and they were found to outperform maximum likelihood and linear discriminant function classifiers. They were used because of their simple, intuitive and explicit classification structure. [25]SVMs have also been used to classify land cover using satellite images and were compared against Neural Network Classifiers, Decision Tree classifiers and Maximum Likelihood classifiers. With proper configuration of the Kernel, they were found to be more accurate than the rest. [26]

4.3 Face Recognition and detection

Face recognition and detection are another critical application which makes use of classifiers.A system was developed that combines local image sampling, self-organizing Map Neural Network to provide dimensionality reduction and invariance to minor changes in images. The convolutional neural network which provides partial invariance to rotation and scaling was used and it extracted successively large features in a hierarchical set of layer. A set of 400 images of 40 individuals was used to analyse the method. [27] SVMs incorporated with binary tree recognition strategy are also used to solve the multi-class face recognition problem. In this approach, the performance of SVM is compared against the eigenface approach with two face databases. [28]   CNN cascades are used for face detection which helps in tackling visual variations due to pose, lighting, expressions and provides powerful discrimination capability and high performance. CNN cascades act on multiple resolutions. Background regions are eliminated in the fast low-resolution stages and only the challenging candidates are evaluated in the high-resolution stages. [29]

4.4 Object Detection

This approach can be extended to multiple objects as well. This category uses a type of image representation called "Integral Image" to compute the features in a fast manner and, consequently, is fed into the detector and a learning algorithm based on AdaBoost to select few critical features. More complex classifiers are combined increasingly in a "cascade" which allows background images to be quickly discarded and thus more computation time can be spent on object-like regions. This can be used in real-time systems. [30] Lastly, ImageNet also used this classification approach. Deep CNN was employed to classify 1.3 million high-resolution images (LSVRC-2010 ImageNet training set) into 1000 different classes. The neural network was composed of 60 million parameters, 500,000 neurons, five Convolutional layers, and 2 globally connected layers with a final 1000 way softmax. Top-1 and Top-5 error rates of 39.7% and 18.9% were achieved, which are considered better than the previous state of the art results. [31]

5. Conclusion

Detection and recognition become an important topic for research. Various algorithms have been proposed for this purpose. This paper attempts to study and provides a brief knowledge about the different image classification approaches. This survey also gives a theoretical knowledge about different classification methods and provides the advantages and disadvantages of various classification methods. This also elucidates the applications of image classification giving examples of already existing systems.

References

[1] B. Boser, I. Guyon and V. Vapnik, "A training algorithm for optimal margin classifiers", Proceedings of the fifth annual workshop on Computational learning theory - COLT '92, 1992.

[2] W.H.Press,   S.A.Teukolsky,   W.T.Vetterling, B.P.Flannery,Section 16.5. Support Vector Machines, Numerical Recipes: The Art of Scientific Computing, ed. 3, Cambridge University Press, 2007.

[3] "Support vector machine", Wikipedia, 2016. [Online]. Available: https://en.wikipedia.org/wiki/Support_vector_machine. [Accessed: 22- Mar- 2016].

[4] "Decision tree learning", Wikipedia, 2016. [Online]. Available: https://en.wikipedia.org/wiki/Decision_tree_learning. [Accessed: 22- Mar- 2016]

[5] S. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology", IEEE Transactions on Systems, Man, and Cybernetics, vol. 21, no. 3, pp. 660-674, 1991.

[6] "Pruning (decision trees)", Wikipedia, 2016. [Online]. Available: https://en.wikipedia.org/wiki/Pruning_(decision_trees). [Accessed: 22- Mar- 2016].

[7] Shepherd, B. A., "An appraisal of a Decision Tree approach to Image Classification". In Proc. of the Eighth

International Joint Conference on Artificial Intelligence, pp. 473–475, 1983.

[8] L. Breiman, "Random Forests", Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[9] T.K. Ho, "Random Decision Forests," Third Int'l Conf. Document Analysis and recognition, pp. 278-282, Montreal, 14-16 Aug. 1995.

[10] E. Kleinberg, "An overtraining-resistant stochastic modeling method for pattern recognition", The Annals of Statistics, vol. 24, no. 6, pp. 2319-2349, 1996.

[11] E. Kleinberg, "On the algorithmic implementation of stochastic discrimination", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 5, pp. 473-490, 2000.

[12] Kleinberg, Eugine. "Stochastic Discrimination and its Implementation"

[13] "K-nearest neighbors algorithm", Wikipedia, 2016. [Online]. Available: https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm. [Accessed: 22- Mar-2016].

[14] N. Altman, "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression", The American Statistician, vol. 46, no. 3, pp. 175-185, 1992.

[15] "Deep learning", Wikipedia, 2016. [Online]. Available: https://en.wikipedia.org/wiki/Deep_learning#Deep_neural_networks. [Accessed: 22- Mar- 2016].

[16] Y. Bengio, "Learning Deep Architectures for AI", FNT in Machine Learning, vol. 2, no. 1, pp. 1-127, 2009.

[17] H. Larochelle, Y. Bengio, J. Louradour and P. Lamblin, "Exploring Strategies for Training Deep Neural Networks", The Journal of Machine Learning Research, vol. 10, pp. 1-40, 2009.

[18] Q. V. Le, J. Ngiam, A. Coates, A. Lahiri, B. Prochnow and A. Y. Ng, "On optimization methods for deep learning", Proc. 28th Int. Conf. Machine Learning, pp. 265-272

[19] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition", Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, 1998.

[20] P. Y. Simard, D. Steinkraus, and J. C. Platt." Best practices for convolutional neural networks applied to visual document analysis". In Seventh International Conference on Document Analysis and Recognition, pp.958-963, 2003.

[21] "Convolutional neural network", Wikipedia, 2016. [Online]. Available: https://en.wikipedia.org/wiki/Convolutional_neural_network. [Accessed: 22- Mar- 2016].

[22] B. Ko, S. Kim and J. Nam, "X-ray Image Classification Using Random Forests with Local Wavelet-Based CS-Local Binary Patterns", Journal of Digital Imaging, vol. 24, no. 6, pp. 1141-1151, 2011.

[23] Y.Fan, D.Shen, C.Davatzikos." Classification of structural images via high-dimensional image warping, robust feature extraction, and SVM." Med Image ComputComput Assist IntervIntConf pp.1-8, 2005

[24] S. Kloppel, C. Stonnington, C. Chu, B. Draganski, R. Scahill, J. Rohrer, N. Fox, C. Jack, J. Ashburner and R. Frackowiak, "Automatic classification of MR scans in Alzheimer's disease", Brain, vol. 131, no. 3, pp. 681-689, 2008.

[25] M. Friedl and C. Brodley, "Decision tree classification of land cover from remotely sensed data", Remote Sensing of Environment, vol. 61, no. 3, pp. 399-409, 1997.

[26] C. Huang, L. Davis and J. Townshend, "An assessment of support vector machines for land cover classification", International Journal of Remote Sensing, vol. 23, no. 4, pp. 725-749, 2002.

[27] S. Lawrence, C. Giles, Ah Chung Tsoi and A. Back, "Face recognition: a convolutional neuralnetwork approach", IEEE Trans. Neural Netw., vol. 8, no. 1, pp. 98-113, 1997.

[28] G. Guo , S. Li and K. Chan, "Support vector machines for face recognition", Image Vis. Comput., vol. 19, no. 9-10, pp. 631-638, 2001

[29] H. Li, Z. Lin, X. Shen, J. Brandt, G. Hua," A convolutional neural network cascade for face detection.", In CVPR, 2015

[30] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", CVPR, 2001

[31] A. Krizhevsky, I. Sutskever, G. Hinton. "ImageNet classification with deep convolutional neural networks". In NIPS, 2012.

Author Profile

Vishal. T. V is an undergraduate student in S.S.N. College of Engineering currently pursuing his third year of B.E. in Computer Science and Engineering. His research interests are primarily Deep Learning and Artificial Intelligence.

Srinidhi. S, is an undergraduate student at SSN College of Engineering, majoring in Information Technology. Her research interests lie in the field of machine learning and data mining.

Srividhya. S, is an undergraduate student at SSN College of Engineering, majoring in Information Technology. Her research interests lie in the field of machine learning and Computer Vision

Sri Vishnu Kumar. K, an undergrad student at SSN College of Engineering, majoring in IT, is a coding enthusiast and a competitive coder. His research interests are machine learning and artificial intelligence.

Swathika. R, Assistant Professor in the Department of Information Technology has 15 years of teaching experience. She received her B.E (CSE) degree with first class from Bharathiyar University, M.Tech (Computer Science and Engineering) degree first class with distinction from National Institute of Technology (NIT), Trichy. She is pursuing Ph.D. (part-time programme) at Anna University, Chennai in the area of Image Processing. Her research interests include Image processing, Feature Extraction and Satellite image analysis.