Capsule Network on Font Style Classification

Dr. T. Vijayakumar,

Professor,

Department of ECE,

GNIT, Hyderabad, India.

Email id: vishal_16278@yahoo.co.in

Mr. R. Vinothkanna,

Department of ECE, Vivekanandha College of Technology for Women,

Namakal, India.

Email: rvinothkannaphd@gmail.com

Abstract: Verification of font style followed in a file is a difficult task to classify. An artificial intelligence based algorithm network can effectively perform this task in reduced time. Capsule network is one among such algorithm and an emerging technique implemented for so many classification process with limited datasets. The proposed font style classification algorithm is enforced with Capsule Network (CapsNet) algorithm for executing the font style classification task. The proposed method is confirmed by classifying times new roman, Arial black and Algerian font style in English letters along with the performance evaluation in terms of accuracy and confusion matrix parameters. The proposed network structure is also compared with the existing Naive Bayes (NB), Decision Tree (DT) and K nearest neighbor (KNN) algorithms for comparative study and the evaluation result indicates that the proposed font style classification model based on CapsNet is classifying the images with better accuracy, F1 score and Gmean.

1. Introduction

Font style classification is a necessary tool for very large file proofing process. Following a particular font throughout a file or paper writing improves the legibility while reading a paper. The font style verification comes under proof reading stage of the material to be printed. Most of the time, proofing of a document on their font size and style are verified by an experienced manual grader to improve the document reading

64

experience. There are chances for failure in manual proofing due to careless mistake and tiredness of human eye. Figure 1 specifies the stages involved in the existing printing process of a document or book. The primary stage consists of literature, writing and typing along with organizing the structure of the document and the secondary stage consists of proofing and printing of the document in the prescribed formant.

Figure 1. Printed document stages

As of now many printing programs are made without computerized proofing process. Very few printing programs are equipped with grammatical error identification and spelling mistake identification tool alone. Figure 2 indicates a sample tool identifies on spelling mistake error, grammatical error and contextual spelling error. The sample tool image also indicates that the tool is not equipped to identify the font style change in the word 'going' from times new roman to arial.
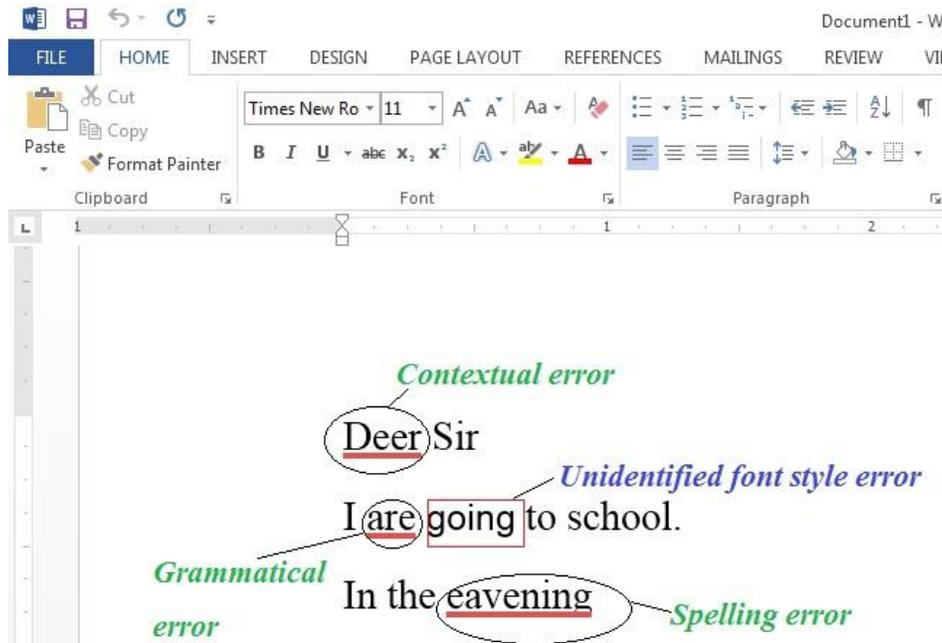

Figure 2. Existing proofing tool performance.

A computer assisted tool can improvise the proofing process including the font style change detection in the document. The reliability of the proofing process can also improve with lesser time when it is computerized. The computerized tool can also work continuously with any number of page in a document

without tiredness or stress as like in manual verification process. This improves the readers experience in reading the content of document. An artificial intelligence based computer vision algorithm is a solution for such problems to classify a word or letter in different style from a document.

2. Related Work

There are artificial intelligence based classifications algorithms are developed for medical image diagnosis, satellite image analysis, environment condition analysis, traffic analysis etc. The section explores the existing classification applications with their algorithm performance along with existing font and text classification approaches. A multiple instance learning approach [1] was introduced to classify the breast cancer image stages. The approach was derived to overcome the drawbacks in the single instance learning approach. The approach makes a bag to organize the similarities between the images for classification process. The results of the approach indicates a better result over than the single instance approach. A CapsNet based multitask learning methodology [2] was evaluated with text classification performance. The methodology is implemented to group the review comments entered by the people on several online platform. The CapsNet methodology is designed with expectation maximization routing strategy instead of the traditional dynamic routing. The algorithm is engaged with primary capsule for managing the information collected from an image through task based routing. Similarly, the secondary capsule layer is involved with length of the routing task information.

A synergic deep learning technique [3] was proposed to classify the medical images with the help of multiple deep convolution neural network. The techniques was developed to overcome the issue on classifying inter-class and intra-class images in a modality. The image information of each image is saved in a deep convolution network and then the image information are interconnected with their respective deep convolution network for matching their similarities to form a separate and single group network. The performance of the technique did a better job in classifying medical images than the existing algorithms. An image classification technique for detecting the vehicle leakage [4] was framed with mean weight least square twin support vector machine algorithm. The algorithm is added with Bayesian filter for removing the noise presence in the image before forwarding it to the classification process. The image features are extracted using multiple genetic algorithm for easy understanding by the classifier. The extracted features are organized to identify the location on leakage occurrence in a vehicle. An X-ray image defect detection system based on deep learning technique [5] was presented to identify the quality of the aluminum parts in an automobile system. The technique is developed to overcome the traditional image processing techniques for better accuracy rate and time saving purpose. Feature pyramid network algorithm is implemented here to train the models for doing the classification part in the process and it is improved in accuracy with 40.9% than the faster R-CNN algorithm. The accuracy of the classification process is also improved than the faster R-CNN.

A fetal echocardiography monitor system based on granular model soft computing technique [6] was structured to improve the performance of the existing intelligent system. Here the features from the echo image is extracted in to several parameters for training and the extracted parameters are estimated through granular model for classification. The fuzzy classifier is added here for performing the classification task from the value generated through granular model and the extracted features in the training model. The experimental results indicates an acceptable improvement in predication accuracy and detection accuracy. A convolution neural network is employed for classifying the echocardiography images [7] in real time in to seven different views for improving the work speed in the medical practice. A convolution block with different values are added here to segregate the images in to different views and a batch normalization process is employed to forma a group of structures in creating the images for various viewing angle. An activation pooling layer is connected in the final layer to optimize the data required for the output layer to the end user for medical verification. The RGBD analysis [8] is carried to isolate the variety of fruits from a group for making the work in to intelligent process. A capsule network algorithm is implemented for splitting the variety of fruits with respect to their color in nature. A depth filter is added further to the network structure for extracting the information lies in the part of image zone for moving it to fuzzy inference system. A density based clustering approach is added after the feature extraction for combining the features to classify the images with a help of LSVM classifier.

An olive fruit variety classification scheme [9] is proposed by merging an image processing technique with convolution neural network algorithm. The fruit image is convoluted in to different segments for making a convolution layer for classifying the images. The image processing technique comprises of HSV transformation and binarization based on Ostu's method along with a morphological operation for noise removal process. Further the image is transferred to marker controlled watershed segmentation scheme for extracting the borders of each olive fruit in an image. Again a morphological operation is insisted to border component removal in an image. The final size and shape of the morphological operated images is marked with respect to their variety for marking them in classification algorithm. A performance analysis on several neural networks on tumor investigation and cancer prediction [10] is carried to identify a better classification algorithm for medical diagnosis. The work is analyzed with feed forward neural network, recurrent neural network and convolution neural network for investigation purposes. All the neural networks are performed with a similar data set for analysis purpose. The performance is moved with an image preprocessing state for noise removal and filtering process for improving the feature predictions from an image. The clustering process divides the images in to several group for sampling the images for classification. The performance result indicates that the convolution neural network is doing a better work with accuracy of classification than the other neural networks.

Convolution neural network based architecture [11] was designed to classify the Chinese letter styles. The deep learning based technique is designed to overcome the drawbacks in the traditional image processing tools. But the designed techniques was able to classify only four different styles in the Chinese letters, where in the traditional method it was able to classify five different Chinese fonts. The added advantage

67

in the neural network based architecture is to collect the unwanted feature information from an image to suppress in a special layer to vomit in the testing process of classification. This improves the performance parametric over than the traditional method in terms of accuracy. An image processing algorithm for smart text recognition [12] was proposed to extract the information on the letters for vocalization to the visually challenged people. The algorithm is equipped with Latte Panda Alpha system to visualize an image and it is forwarded to an image processing algorithm for font segmentation and feature extraction of an image. A text speech synthesizer is added to the system for reading the extracted features in to vocalization process. The system attains 97% of accuracy in assisting the visually challenged people. The accuracy can be improved when the written information is clear and legible. An improved naïve Bayesian algorithm [13] was developed to classify the text image applications. The text image information are extracted with respect to their particularity and the semantic information extracted from the images are categorized into several layers for comparison. The proposed classification system is improved to overcome the downsides of the traditional navie Bayesian algorithm. A mutual information method is newly added to the prescribed system to verify the correlation between the extracted information of the training data and the current data. The experimental result indicates that the improved technique works better than the previous state of art work.

A biometric recognition system based on CapsNet [14] architecture is developed for personal identification of a person which is needed in many applications. The CapsNet algorithm is permitted here to overcome the shortcomings of the convolution neural network. The deep learning algorithms are generally independent in learning the features from an image, but due to presence of noise in an image the efficiency may reduce to certain level. The CapsNet has the ability to learn the features with comparatively lesser samples required for convolution neural network and the ability to learn the features from an image is also excellent than the convolution neural network. A Gabor wavelet is connected to the system for enhancing several features from an image for better extraction of information by the network algorithm. The CapsNet has the ability to extract the max pooling information from an image along with the vector outputs [15]. This makes the CapsNet to collect additional information about an image more precisely with respect to their location than the convolution neural network algorithms. A CapsNet based text classification algorithm [16] was framed to minimize the computational complexity of the deep learning algorithm in classification. The computational complexity is reduced with a help of dynamic routing along with the convolution layer information. The algorithm is further modified the traditional dynamic routing in the CapsNet in to static routing for minimizing the computation with maximum accuracy.

3. Proposed Work

The proposed CapsNet based font classification algorithm is designed to catch out the font style changes in a document or a phrase. This makes an intelligent proofing system for document verification process. The proposed model is verified with a self-made dataset consists of English capital letters of times news roman, Arial black and Algerian with different font sizes. The dataset consists of 702 font images, in that 468 are separated as training dataset and 234 are mode as testing dataset. Figure 3 indicates few font images from the dataset. The proposed model performance is analyzed with certain existing modalities for verifying the efficiency of the proposed model.
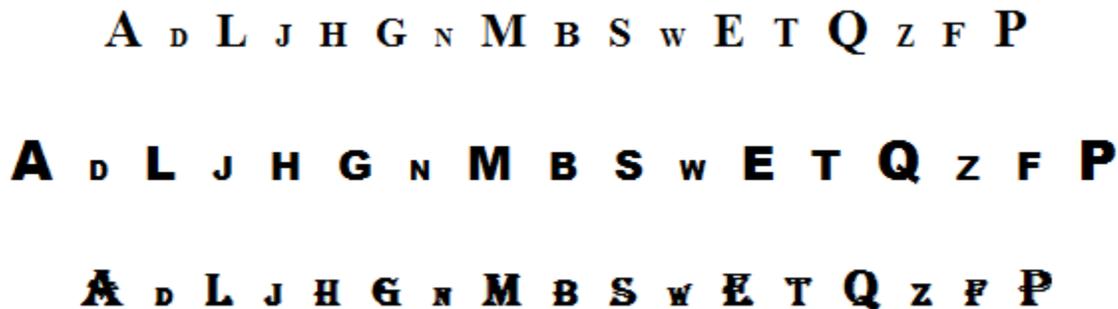

Figure 3. Dataset images.

The images are preprocessed for better training process to the classifier system. The letters in the document or phrase are converted in to uppercase for analysis and then the letters in the words are separated in to an individual one for analyzing process. The process of disconnecting the words in to separate letters are called are tokenization. The tokenized images are gone through boundary estimation for finding the shape of a letter. Then the letter images are transferred to different classifiers for analyzing its performance over than the proposed CapsNet algorithm. Figure 4 indicates the structural flow of the proposed text classification system. The preprocessed images are trained to the module for extracting the necessary information from the font images regarding their shape and the training sample data information are marked as model data for comparison purpose. The classifier is the proposed algorithm module able to read the information stored on the model data for comparing it with the preprocessed testing data. The testing data is also preprocessed as same as like of the training data for making the edge of the fonts very clear to the classifier for better absorption. The classifier classifies the images with respect to their nature of algorithm and performs the process with various accuracy with respect to time and testing data information.
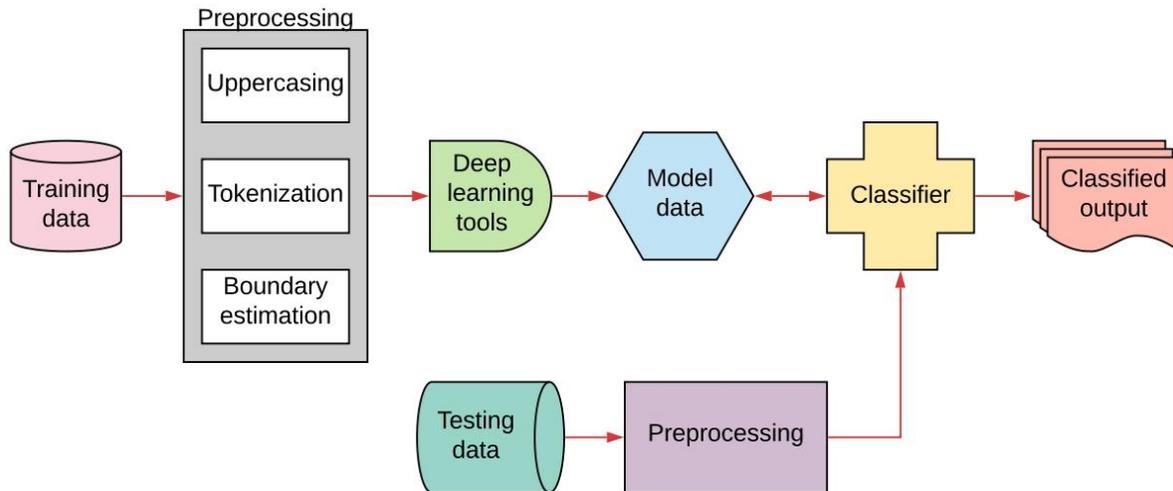
69

Figure 4. Flow of the proposed font classification process.

The naive Bayesian algorithm is widely applied to several text classification models for its acceptable accuracy in large data training modules. The naïve Bayesian network algorithms are probabilistic graphical network to trace out the information of the data edges in distribution E(X). The probability details are analyzed inside the classifier with respect to the conditions of the testing data information. The Bayesian formula is utilized here for analyzing the probability status of the testing data. The testing data are classified to where the maximum probability is attained by the classifier algorithm. The Bayesian probability is calculated with respect to the following equation E (H|X).

$$E(H|X) = \frac{E(X|H)\,E(H)}{E(x)}$$

The E (H) is the hypothesis probability of a preliminary data collected from the sample images used for training the classifiers. The probability of the training data X is expressed as E (X). E(H|X) is the post operated probability of the data at condition X and the E(X|H) is the probability state of the X at condition H.

The decision tree algorithm is also taken in to account for performance comparison of the proposed algorithm. The decision tree classifies the images with respect to their primary class and further the images are separated with respect to the class in which it belongs to. The information and features extracted from the training data are made in to several nodes and the nodes are interconnected with several multiple nodes for saving the collected data information. The classifier takes the testing data and

70

verify its similarity to the model data in the tree structure hierarchy to find out its nature node selection. The decision tree algorithm is a complex in computation of results as well as in training. It takes more time for training the data to match the features to the existing tree nodes. But the output of the decision tree algorithms are generally high in accuracy. Figure 5. Shows how the nodes in the decision tree algorithm is connected to the classifier. The decision tree takes more time for training but the execution speed is high for testing a data. The decision tree algorithms have the ability to work with noisy data for classification process and it is very effective when there is large data is need to be classified.
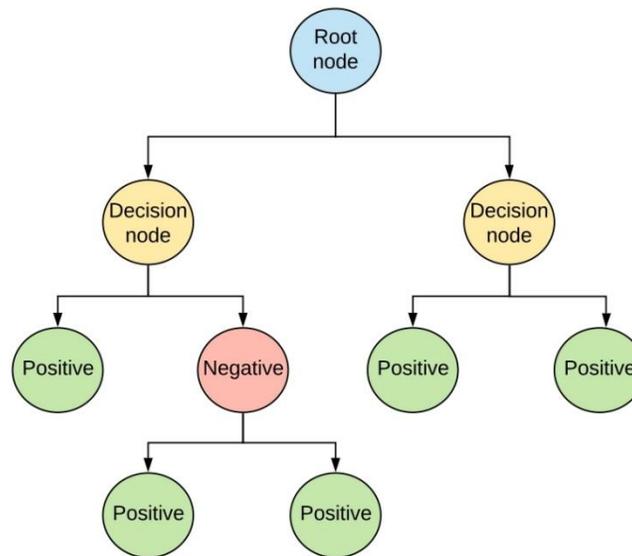

Figure 5. Decision tree algorithm pattern.

The KNN algorithm extract the features from an image and marks it as K value in the model data for comparing the features with the testing data. The K value is calculated for each testing data to verify its similarity to the model data. The similarity distance is marked for indexing the feature values to sort it out to the existing training classification closeness. The closer value of K makes the testing data into specific classification of images. The model data mentioned in the algorithm is generally called as training sample labels in the KNN algorithm. The features of the labeled value makes a tree kind of structure for decision making in the testing process. The distance between the K values are calculated as follows.

$$G(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

The equation works for n number of variations in the feature extraction for labeling the K value. The nearest value of distance makes the test image belongs to a specific class. The location information is labeled as G(x,y) for comparing the same location pixels with the training label information.

The CapsNet is a deep learning algorithm designed to collect the information from an image efficiently without losing it. The image information are lost in case of the traditional convolution neural network due to high density in the pooling layer. The CapsNet was designed to overcome such drawbacks in the deep learning algorithms. The CapsNet algorithms have the ability to train a model data with limited sample when comparing with any other deep learning algorithm. The information available in the images are collected through convolution1d layer along with primary caps and digit caps for operations. The CapsNet algorithms are framed with large number of hidden neurons for gathering the labeled information in to specific data location for future classification process. The probability of the classified images are conducted with the help of a non-linear squashing function with output vector E_V. The iterative dynamic routing is determined by the coupling coefficient of the capsules and the length of the capsules are maintained by a weight matrix. The length modified matrix functions are moved towards the lower layer of the capsule networks.

$$E_v = \frac{||G_V||^2}{\rho + ||G_V||^2} \frac{G_V}{||G_V||'}$$

Where

E_v is the output vector of the CapsNet and G_v is the input vector of the CapsNet.

ρ is the parameter derived to customize the length of the vector capsules.

The input vector of the capsules are expressed as the product of the output vectors u_v with matrix weight of W_{uv}. C_{uv} is the coupling coefficient identified by the routing process of the capsules. The training and test capsules probability log is mentioned as b_{uv} and b_{ux}.

$$G_v = \sum_i C_{uv} W_{uv} u_v$$

$$C_{uv} = \frac{\exp b_{uv}}{\sum_k \exp b'_{ux}}$$

4. Results and Discussion

The proposed CapsNet based font style classification algorithm performance is compared with naïve Bayesian algorithm, decision tree and K nearest neighbor algorithm. The dataset images are commonly distributed for training and testing in all the algorithms. The simulation result is acquired from an i7 processor system with 8GB RAM configuration. The overall accuracy, F1 score and Gmean are taken into account for analyzing the performance of the classification algorithms with the proposed model. F1 score detects the harmonic sensitivity mean and Positive Predictive Value (PPV) together at the same dimension. The Geometric mean (Gmean) is calculated from the sensitivity and specificity of the algorithms. Table 1 indicates the results observed by the algorithms on their performance. Figure 6 and 7

Artificial Intelligence
&
Capsule Networks

explores the comparison analysis of the algorithms. The expression for F1 score and Gmean is expressed in percentage as follows,

$$F1\ score = \frac{2\ (\alpha)(\rho)}{(\alpha + \rho)}$$
$$Gmean = \sqrt{\alpha \cdot \beta}$$

Where,

α and β are the sensitivity and specificity of the classified results observed from their confusion matrix results ρ is the PPV value.

Table 1. Result performance of algorithms.

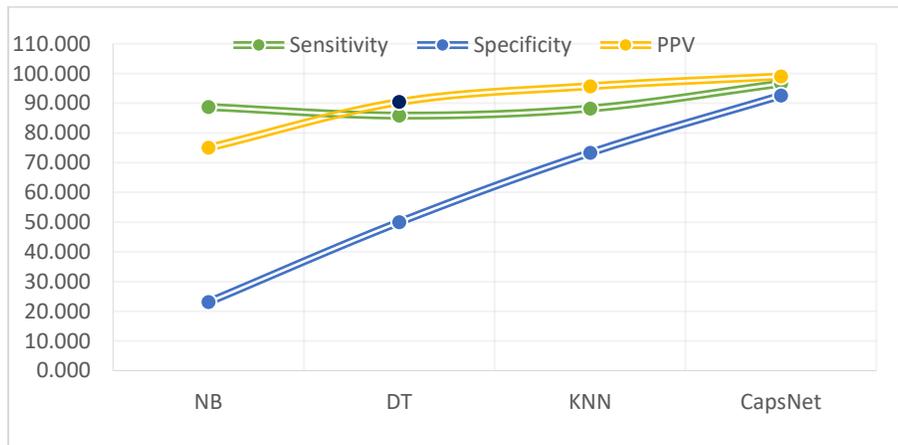| Algorithm	Accuracy	Sensitivity (α)	Specificity (β)	PPV (ρ)	F1 Score	Gmean
NB	70.512	88.757	23.076	75.000	81.301	45.258
DT	80.341	85.858	50	90.426	88.083	65.520
KNN	86.324	88.235	73.333	95.745	91.837	80.440
CapsNet	96.153	96.618	92.592	99.010	97.800	94.584

Figure 6. Performance comparison in terms of sensitivity, specificity and PPV
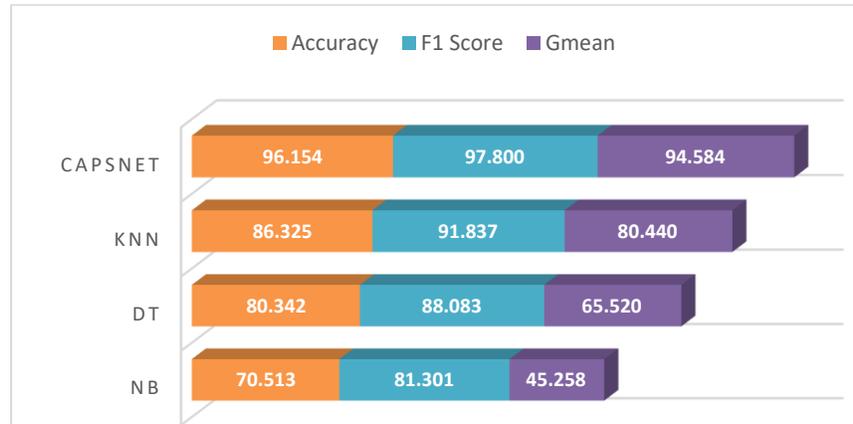
Figure 7. Performance comparison in terms of accuracy, F1score and Gmean.

The performance evaluation figures indicates that the proposed CapsNet based font style classification algorithm performs better than the existing algorithms. The accuracy rate, F1 score and Gmean parameter proves that the proposed CapsNet based algorithm has the ability to classify the images with lesser training data. The naïve Bayes algorithm has the ability to perform faster than the KNN algorithm but the accuracy rate is reduced because of lesser training data. The performance of naïve Bayes algorithm can be improved once the number of training data is increased. The training images need not to be preprocessed in naïve Bayes algorithm, as it has the nature of dealing with noisy images. But in the proposed performance evaluation process has the preprocessing steps for dealing with other algorithms. Hence, the training images transferred to naïve Bayes algorithm is also preprocessed.

The decision tree algorithm is also have the ability to classify the images with higher accuracy when the training data is huge amount. But the accuracy of the decision tree algorithm is far better than the naïve Bayes algorithm. The KNN algorithm performance is comparatively better than the NB and DT algorithm because of its nature that it can classify the data with lesser training sample. The CapsNet is also similar to the KNN works well with the smaller dataset. Along with that advantage the CapsNet can collect more information from an image for labeling process with lesser network architecture. The performance speed of the KNN is reduced when the number of neighbor values is increased respect to the image information. The CapsNet overcomes such limitations on processing the data labels of an image.

5. Conclusion and Future Scope

The proposed font style classification model is performed with different classifiers for identifying a suitable algorithm for the process. Hence several algorithms are performed with the same dataset for the comparison analysis. The comparison performance result indicates that the CapsNet based algorithm is performing better than the other algorithms in terms of accuracy with 96.15%, F1 score of 97.8% and geometric mean of 94.58%. The performance flow is carried with a preprocessing step for noise removal and boundary segmentation for improving the learning ability of the algorithms. The proposed algorithm structure can be improved and verified by removing the preprocessing step and include with large amount of training data in the future for verifying the performance changes in the performed algorithms.

References

[1] Sudharshan, P. J., Caroline Petitjean, Fabio Spanhol, Luiz Eduardo Oliveira, Laurent Heutte, and Paul Honeine. "Multiple instance learning for histopathological breast cancer image classification." *Expert Systems with Applications* 117 (2019): 103-111.

[2] Jacob, I. Jeena. "Performance Evaluation of Caps-Net Based Multitask Learning Architecture for Text Classification." *Journal of Artificial Intelligence* 2, no. 01 (2020): 1-10.

[3] Zhang, Jianpeng, Yutong Xie, Qi Wu, and Yong Xia. "Medical image classification using synergic deep learning." *Medical image analysis* 54 (2019): 10-19.

[4] Manoharan, Samuel. "Image Detection Classification and Recognition for Leak Detection in Automobiles." *Journal of Innovative Image Processing (JIIP)* 1, no. 02 (2019): 61-70.

[5] Du, Wangzhe, Hongyao Shen, Jianzhong Fu, Ge Zhang, and Quan He. "Approaches for improvement of the X-ray image defect detection of automobile casting aluminum parts based on deep learning." *NDT & E International* 107 (2019): 102144.

[6] Sathesh, A. "Performance analysis of granular computing model in soft computing paradigm for monitoring of fetal echocardiography." *Journal of Soft Computing Paradigm (JSCP)* 1, no. 01 (2019): 14-23.

[7] Østvik, Andreas, Erik Smistad, Svein Arne Aase, Bjørn Olav Haugen, and Lasse Lovstakken. "Real-time standard view classification in transthoracic echocardiography using convolutional neural networks." *Ultrasound in medicine & biology* 45, no. 2 (2019): 374-384.

[8] Chandy, Abraham. "RGBD Analysis for Finding the Different Stages of Maturity of Fruits in Farming." *Journal of Innovative Image Processing (JIIP)* 1, no. 02 (2019): 111-121.

[9] Ponce, Juan M., Arturo Aquino, and José M. Andújar. "Olive-Fruit Variety Classification by Means of Image Processing and Convolutional Neural Networks." *IEEE Access* 7 (2019): 147629-147641.

[10]   Vijayakumar, T. "Neural network analysis for tumor investigation and cancer prediction." *Journal of Electronics* 1, no. 02 (2019): 89-98.

[11]   Zhang, Jiulong, Mingtao Guo, and Jianping Fan. "A novel CNN structure for fine-grained classification of Chinese calligraphy styles." *International Journal on Document Analysis and Recognition (IJDAR)* 22, no. 2 (2019): 177-188.

[12]   Manoharan, Samuel. "A smart image processing algorithm for text recognition information extraction and vocalization for the visually challenged." *Journal of Innovative Image Processing (JIIP)* 1, no. 01 (2019): 31-38.

[13]   Gao, Hongyi, Xi Zeng, and Chunhua Yao. "Application of improved distributed naive Bayesian algorithms in text classification." *The Journal of Supercomputing* 75, no. 9 (2019): 5831-5847.

[14]   Jacob, I. Jeena. "Capsule network based biometric recognition system." *Journal of Artificial Intelligence* 1, no. 02 (2019): 83-94.

[15]   Smys, S., Joy Iong-Zong Chen, Shahed Mohammadi, and Álvaro Rocha. "Capsule networks and autonomous systems." *International Journal of Intelligent Unmanned Systems* (2019).

[16]   Kim, Jaeyoung, Sion Jang, Eunjeong Park, and Sungchul Choi. "Text classification using capsules." *Neurocomputing* 376 (2020): 214-221.

Authors Biography

Dr. T. Vijayakumar, works as a Professor, in the Department of electronics and communication in GNIT, Hyderabad, India. His interested area of research includes capsule networks , Machine Perception, Robotics, Cyber-Physical Systems, Internet of Things, Complex Networks, Quantitative Network-based modelling.

Mr. R. Vinothkanna, works in the Depart ment of ECE in Department of ECE, Vivekanandha College of Technology for Women, Namakal, India. The author's area of interest includes Complex and Intelligent Systems, Networks, Recommendation System, Human-Computer Interface, and Knowledge Representation