Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 167 (2020) 2554–2560

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Computational Intelligence and Data Science (ICCIDS 2019)

Hybrid CNN-SVM Classifier for Handwritten Digit Recognition

Savita Ahlawat[a], Amit Choudhary[b]*

[a]Reader, Deptt. of Computer Science and Engineering, Maharaja Surajmal Institute of Technology, New Delhi
[b]Associate Professor, Deptt. of Computer Science, Maharaja Surajmal Institute, New Delhi

Abstract

The aim of this paper is to develop a hybrid model of a powerful Convolutional Neural Networks (CNN) and Support Vector Machine (SVM) for recognition of handwritten digit from MNIST dataset. The proposed hybrid model combines the key properties of both the classifiers. In the proposed hybrid model, CNN works as an automatic feature extractor and SVM works as a binary classifier. The MNIST dataset of handwritten digits is used for training and testing the algorithm adopted in the proposed model. The MNIST dataset consists of handwritten digits images which are diverse and highly distorted. The receptive field of CNN helps in automatically extracting the most distinguishable features from these handwritten digits. The experimental results demonstrate the effectiveness of the proposed framework by achieving a recognition accuracy of 99.28% over MNIST handwritten digits dataset.

Peer-review under responsibility of the scientific committee of the International Conference on Computational Intelligence and Data Science (ICCIDS 2019).

Keywords: Handwritten digit recognition; Hybrid model; Convolutional Neural Network; Support Vector Machine.

1. Introduction

Now a days, Handwritten Digit recognition is an active area of research in the domain of handwriting recognition. In recent years, many handwritten digit recognition systems have been proposed for practical applications which demand high recognition accuracy and reliability. The touched and overlapped characters, different handwriting

* Corresponding author. Tel.: +91-9911335069
E-mail address: amit.choudhary69@gmail.com

1877-0509 © 2020 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of the scientific committee of the International Conference on Computational Intelligence and Data Science (ICCIDS 2019).
10.1016/j.procs.2020.03.309

patterns and handwriting styles of individuals adds to the complexity of the recognition system. In the area of handwriting recognition, several methods have been proposed in the literature such as Artificial Neural Network (ANN), Neuro-Fuzzy Systems (NFS), Support Vector Machine (SVM) and deep learning-based classifiers [1-9]. Although decent recognition accuracy has been reported by these classifiers; handwriting digit recognition is still an open research problem and demands for exploring new techniques and methodologies that would further improve the performance in terms of recognition accuracy, running time and computational complexity. Therefore, in this work, a framework of hybrid CNN-SVM is proposed for handwritten digit recognition. The focus of the paper is to extract the features from the input handwritten digit images of MNIST dataset using CNN. These learned features are then passed to the SVM classifier for the proposed handwritten digits recognition experiment.

A lot of research work based on extracting handcrafted features has been reported in the literature. This manually feature extraction method is a tedious process and needs human expertise. Also, the handcrafted feature extraction method involves a trade-off between efficiency and recognition accuracy because the processing of extraneous features may increase computation overhead leading to the poor performance of the recognition system. On the other hand, non-handcrafted features extraction process comprises of retrieving features directly from the raw images. This method eliminates the necessity of collecting prior knowledge and design details of features.

The major advantage of using the CNN model is that it exploits the topological information present in the input and is invariant to basic transformations like rotation, translation etc. On the other hand, MLP models never considers detailed topology information of input and are not suitable for complex problems. The organization of the rest of the paper is as follow. Section 2 presents related work done so far by the researchers in this field. Section 3 describes the proposed methodology. Experimental setup and discussion of results are presented in section 4 and the conclusion and future scope are presented in section 5.

2. Related Work

A significant number of research papers in the field of handwriting recognition are available in the literature [10-20]. In 1995, SVM has been used for the first time for handwritten digit OCR [21]. Later, SVM classifiers have emerged as a default choice for various supervised classification problems like character recognition [17-19] face detection [22-23] and object recognition [27-30]. Boukharouba and Bennia proposed a Freeman Chain Code approach for feature extraction and developed a Support Vector Machine based handwritten digit recognition system [31].

In recent years, the convolutional neural networks have been effectively used for handwritten digit recognition and primarily for benchmark MNIST handwritten digit dataset. Most of the experiments achieved high recognition accuracy more than 98% or 99% [32]. The high recognition accuracy of 99.73% on MNIST dataset is achieved while experimenting with the famous committee technique of combining multiple CNN's in an ensemble network [33]. The work was further extended into 35-net committee from the earlier 7-net committee and reported very high accuracy of 99.77% [34]. The Deep Belief Networks (DBN) with three layers along with a greedy algorithm were investigated for MNIST dataset and reported the accuracy of 98.75% [35]. The bend directional feature maps were investigated using CNN for in–air handwritten Chinese character recognition [36]. Lauer et al. work on a LeNet5 convolutional neural network architecture-based feature extractor for the MNIST database [5]. The work reported excellent recognition accuracy. The impressive performance of the research work clearly shows the effectiveness of CNN feature extraction step. The present work is also motivated by the performance of deep learning methods in handwriting recognition domain [37-39].

The structural risk minimization ability of SVM and deep feature extraction ability of CNN when combined together has proved to be extremely useful in many domains [40-45]. The fusion of CNN-SVM can be highly useful in handwriting recognition and hence the aim of the present work. Niu and Suen integrates the CNN and SVM for MNIST digit database and reported a recognition rate of 99.81% [6]. The authors used rejection rules to achieve high reliabilities. Guo et al. investigated hybrid CNN-Hidden Markov Model (HMM) for house numbers recognition from the street view images [46]. The experimental results demonstrated that the hybrid CNN-SVM model effectively improved the recognition accuracy for handwritten digit.

3. Proposed Methodology

Here, a hybrid model of CNN-SVM is proposed for the classification of handwritten digits of MNIST dataset. The proposed system incorporates the best qualities of SVM and CNN classifiers. A convolutional neural network (CNN) consists of multiple fully connected layers and have a supervised learning mechanism. CNN works in a similar fashion as we human do and can learn invariant local features very well. It can extract the most discriminating information from raw digit images. In the proposed system, a 5x5 kernel/filter is used to extract most distinguishable features from the raw input images. In convolutional layer, a $n \times n$ input neurons of the input layer are convoluted with $m \times m$ filter and in return deliver an output of size $(n-m+1) \times (n-m+1)$. The output from each layer becomes the input to the next layer. The receptive field feature of CNN is used to calculate effective sub-regions from the raw digit image.

Support Vector Machine (SVM) aims to represent multi-dimensional dataset in a space where data elements belonging to different classes are separated by a hyperplane. The SVM classifier has the ability to minimize the generalization error on unseen data. The separating hyperplane is also called as an optimal hyperplane. SVM is found to be good for binary classification and is considered poor for noisy data. The shallow architecture of SVM presents some difficulties in learning deep features.

In present work, the hybrid model of CNN-SVM is proposed in which SVM is used as a binary classifier and replaces the softmax layer of CNN. CNN works as a feature extractor and SVM as a binary classifier. The architecture of the proposed hybrid CNN-SVM model is described in Fig. 1.

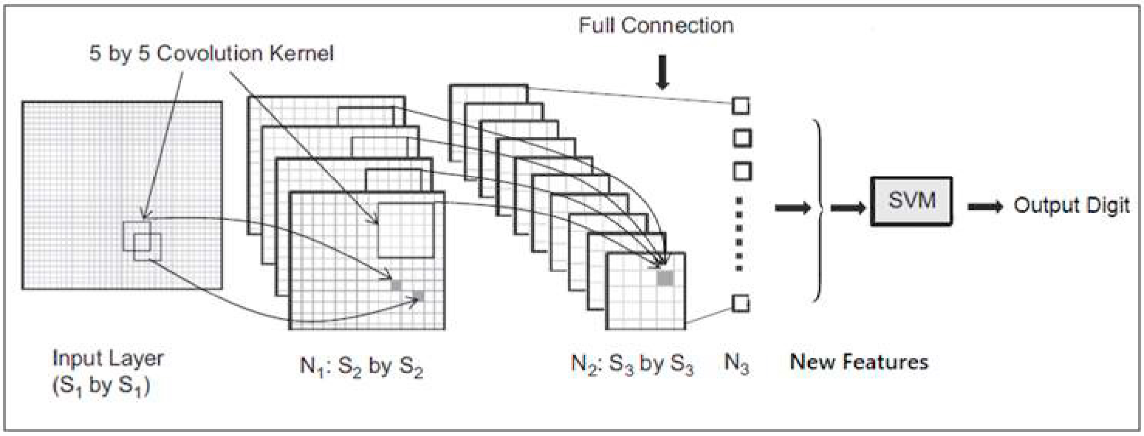

Fig. 1. Architecture of the proposed Hybrid CNN-SVM Model

The model consists of a simple CNN architecture and SVM classifier. The CNN input layer takes a 28×28 matrix of normalized and centered handwritten digit taken from MNIST dataset. A 5×5 convolutional filtering and stride of size 2 are used at the convolutional layers. The N1 and N2 feature map layer extract values which are further considered as distinguishing features of the input image. The CNN is trained after running several epochs and until the training process converges. The last layer of CNN is replaced here by SVM classifier. The features of input handwritten digit obtained in N3 layer are treated as an input for the SVM classifier. The SVM classifier is initially trained with these new automatically generated features of training images. Finally, the trained SVM classifier is used in recognizing the digits used for testing.

4. Experimental Setup and Discussion of Results

The proposed hybrid model is evaluated for MNIST handwritten digits is the proposed experiment. The experimental setup involves the following steps-

4.1. Data Set used

The MNIST a well-known handwritten digit benchmark dataset and has been used in the present work for training and testing the proposed classification algorithm. The dataset consists of handwritten images of the 10 digits (0 to 9). The images are already normalized and present in unsigned byte form. The MNIST dataset is created by taking a subset of a larger dataset available from National Institute of Standards and Technology (NIST) and contains 60000 examples for training and 10000 examples of testing, i.e., a total of 70000 data elements. The images present in the NSIT database were represented using a 20x20 pixel matrix. But the images had some gray levels present in them. This problem has been solved in the MNIST dataset by normalizing all the 20×20 handwritten images into 28×28 images. The center of mass of the earlier image is found out and accordingly, the normalized image is also positioned inside the 28×28 image in such a way that the center of mass is positioned at the center [47]. Fig. 2 shows some sample handwritten digits present in the MNIST dataset. It is also observed that the MNIST dataset is a clean dataset in which no noise is present.

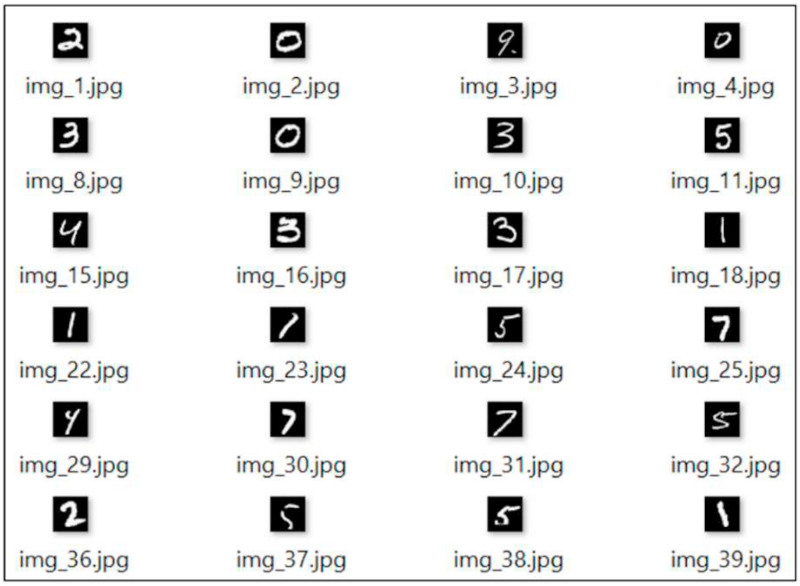

Fig. 2. Sample handwritten digit images from MNIST dataset

4.2. Data Set Preprocessing

The pre-processing is a preliminary step which includes the entire range of steps necessary to bring the input data into a form acceptable for feature extraction. It involves getting character input, noise reduction, centering and scaling of the input image. This is a very crucial process because feature extraction can be misled if the input images are not properly pre-processed. In the present work, the MNIST handwritten digits images are already size normalized and centered.

4.3. Feature Extraction

In proposed work, features from the input handwritten digit images are extracted using the proposed model. A pre-processed handwritten digit image is passed as an input to the proposed hybrid model. The generalization ability of CNN is greatly affected by the three numbers N1, N2 and N3. Therefore, in the present work, N1=50, N2=50 and N3=100 has been taken by considering the previous research [48,49]. The Sigmoid function has been used as an activation function. The function estimates the output probability for each input sample digit image. The linear combination of the outputs from the previous hidden layer with trainable weights and bias term becomes the input for the next layers. Similarly, a feature map is obtained at the third layer i.e. N3 layer. The feature map is converted into a single column vector. This single column feature vector containing distinguishable features is fed to the SVM classifier for recognition experiment.

4.4. Classification

After the pre-processing and feature extraction steps are over, classification of handwritten digit images is carried out by SVM classifier. The training of SVM classifier has been done by taking the feature vectors stored in matrices form. The testing of the digit has been done by using the result of training. Whereas, in hybrid CNN-SVM model, the automatically generated features are passed to the SVM module for training and testing the handwritten digit dataset. The testing data of MNIST is pre-processed in a similar way and then it is used to test the classifier.

The accuracy of the SVM classifier and hybrid SVM-CNN is measured for both training and testing data and the results obtained are noted down. An RBF function is used as a kernel in the hybrid model. The SVM parameters like degree and gamma of the kernel function and the shape of the decision function are chosen carefully as they are the influencing factors during SVM classification.

4.5. Observation and Discussion of Results

The influence of SVM parameters on the classification results of training and testing dataset in both the cases has been presented in Table 1. It can be observed from Table-1 that changing the decision function from one-vs-one to one-vs-rest reduces the accuracy of the classifier. When the value of gamma is 0.1 instead of 1, the accuracy of the classifier is higher, as it can be observed from rows 1 and 4 in Table-1. Increasing the degree results in a reduction in accuracy but there is an odd behavior i.e. the training accuracy increases when the degree is 5 and gamma is 0.1. The testing accuracy decreases in the same manner. The maximum training accuracy is 99.28% for hybrid CNN-SVM when gamma is equal to 0.1, the degree is equal to 5 and the decision function is one-vs-one. The maximum testing accuracy is 98.95% when gamma is equal to 0.1, the degree is equal to 3 and the decision function is one-vs-one.

Table 1. Comparison of Training and Testing accuracy of SVM and CNN-SVM.

Sr. No.	Parameters			Percentage Accuracy (SVM)		Percentage Accuracy (Hybrid CNN-SVM)	
	Gamma	Degree	Decision Function	Training	Testing	Training	Testing
1	1	3	one-vs-one	97.80	97.52	98.80	98.82
2	1	3	one-vs-rest	97.38	97.74	98.48	98.74
3	1	5	one-vs-one	96.18	96.37	97.38	97.39
4	0.1	3	one-vs-one	97.95	97.85	98.95	98.95
5	0.1	3	one-vs-rest	98.93	97.84	98.93	98.84
6	0.1	5	one-vs-one	98.35	97.68	**99.28**	98.88

After analysing the results presented in Table-1, the accuracy of the hybrid CNN-SVM classifier for recognition of handwritten digit is found to be 99.28% which is greater than the recognition accuracy of 98.35% of SVM classifier.

5. Conclusion and Future Scope

In this paper, a hybrid model of CNN-SVM is proposed for handwritten digit recognition that involves automatic feature generation using CNN and output prediction using SVM. The model combines the advantage of CNN and SVM classifiers in recognizing handwritten digits. The model also emphasizes the use of automatically generated features over the hand-designed features. The experimental results showed that our proposed approach achieved the classification accuracy of 99.28% for the MNIST dataset. The research on the hybrid CNN-SVM model is in its early stage and can be further improved. In future, the proposed model can be improved for recognition of handwritten characters in different languages such as French, English, Hindi, Bengali, etc. Some optimizing techniques can also be investigated to boost the classification performance.

References

[1] Larochelle H., Bengio Y., Louradour J. and Lamblin P. (2009) "Exploring Strategies for Training Deep Neural Networks." *Journal of Machine Learning Research* 10:1-40.

[2] Wang, L., Wang, Y. and Chang, Q. (2016) "Feature selection methods for big data bioinformatics: A survey from the search perspective." *Methods* 111:21-31.

[3] Goltsev, A. and Gritsenko, V. (2012), "Investigation of efficient features for image recognition by neural networks", *Neural Networks* 28:15-23.

[4] Kang, M., Palmer-Brown and D. (2008) "A modal learning adaptive function neural network applied to handwritten digit recognition." *Information Sciences—Informatics and Computer Science, Intelligent Systems, Applications* 178(20) :3802-3812.

[5] Lauer F., Suen Y. C. and Bloch G. (2007) "A trainable feature extractor for handwritten digit recognition." *Pattern Recognition* 40(6):1816-1824, 2007.

[6] Niu, X. X. and Suen, C. Y. (2012) "A novel hybrid CNN–SVM classifier for recognizing handwritten digits", *Pattern Recognition* 45(4):1318–1325.

[7] Choudhary, A., Rishi, R., and Ahlawat, S. (2012). "Unconstrained Handwritten Digit OCR Using Projection Profile and Neural Network Approach." In: *Proceedings of International Conference on Information System Design and Intelligent Applications (INDIA 2012), Springer, Heidelberg, 132,* 119-126.

[8] Choudhary, A., Rishi, R. and Ahlawat, S. (2013) "A New Approach to Detect and Extract Characters from Off-Line Printed Images and Text." In: *Proceedings of International Conference on Information Technology and Quantitative Management (ITQM), Procedia Computer Science, 17,* 434–440.

[9] Choudhary, A. and Rishi, R. (2011) "Improving the Character Recognition Efficiency of Feed Forward BP Neural Network." *International Journal of Computer Science & Information Technology (IJCSIT),* 3(1), 85-96.

[10] Graves A. and Schmidhuber J. (2009) "Offline handwriting recognition with multidimensional recurrent neural networks." in: *Advances in neural information processing systems* 545–552.

[11] Sayre (1973) "Machine recognition of handwritten words: A project report." *Pattern recognition* 5 (3):213–228.

[12] Plamondon, R. and Srihari S.N. (2000) "Online and off-line handwriting recognition: a comprehensive survey." *IEEE Transactions on pattern analysis and machine intelligence* 22 (1):63–84.

[13] Yuan, A., Bai, G., Jiao, L. and Liu, Y. (2012) "Offline handwritten English character recognition based on convolutional neural network" in: *Document Analysis Systems (DAS)*, 2012 10th IAPR International Workshop on, IEEE, 125–129.

[14] Manisha, C. N., Reddy, E. S. and Krishna, Y. S. (2016) "Role of offline handwritten character recognition system in various applications" *International Journal of Computer Applications* 135 (2):30–33.

[15] Sa´nchez, J. A., Bosch, V., Romero, V. and K. Depuydt, J. de Does (2014) "Handwritten text recognition for historical documents in the transcriptorium project, in: *Proceedings of the First International Conference on Digital Access to Textual Cultural Heritage, ACM* 111–117.

[16] Pl¨otz, T. and Fink, G.A. (2009) "Markov models for offline handwriting recognition: A survey." *Int. J. Doc. Anal. Recognit.* 12 (4):269–298.

[17] Choudhary A., Ahlawat, S. and Rishi R. (2014) "A Binarization Feature Extraction Approach to OCR: MLP vs. RBF." In: *Proceedings of the International Conference on Distributed Computing and Technology ICDCIT, Springer,* 341-346.

[18] Choudhary, A., Rishi, R. and Ahlawat, S. (2013) "Off-Line Handwritten Character Recognition using Features Extracted from Binarization Technique." In: *Proceedings of International Conference on Intelligent Systems and Control, AASRI Procedia,* 4, 306-312.

[19] Choudhary A. and Rishi R. (2014) "A Fused Feature Extraction Approach to OCR: MLP vs. RBF." In: *Proceedings of the 48th Annual Convention of Computer Society of India- Vol I. Advances in Intelligent Systems and Computing, Springer, Cham*, 248, 159-166.

[20] Choudhary, A. (2014) "A Review of various Character Segmentation Techniques for Cursive Handwritten Words Recognition." *International Journal of Information & Computation Technology (IJICT)* 4 (6), 559-564.

[21] Cortes, C. and Vapnik V. (1995) "Support vector networks", *Machine Learning*, 20

[22] Pontil, M. and Verri, A. (1998) "Support Vector Machines for 3-D Object Recognition." *IEEE Trans. Patt. Anal. Machine Intell*. 20:637-646.

[23] Osuna, E., Freund, R. and Girosi, F. (1997) "Training Support Vector Machines: An Application to Face Detection." *Proc. of CVPR, Puerto Rico. IEEE* 130-136.

[24] Malon, C., Uchida, S. and Suzuki, M. (2008) "Mathematical symbol recognition with support vector machines." *Pattern Recognition Letter*, 29(9):1326–32.

[25] Mahto, M.K., Bhatia, K. and Sharma, R. K. (2015) "Combined horizontal and vertical projection feature extraction technique for Gurmukhi handwritten character recognition." *2015 International Conference on Advances in Computer Engineering and Applications*. 59–65.

[26] Murthy OVR. (2011) "Zoning based Devanagari Character Recognition." *International Journal of Computer Applications* 27(4): 21–25.

[27] Burges, C.J.C. (1998) "A tutorial on support vector machines for pattern recognition." *Data Mining Knowl. Discovery* 2:121–167.

[28] Guo, G., Jain, A.K. and Zhang, H. (2002) "Learning similarity measure for natural image retrieval with relevance feedback." *IEEE Trans. Neural Networks* 13(4):811–819.

[29] Weston, J.A.E. (1999) “Extensions to the support vector method.” Ph.D. Thesis. Royal Holloway College, University of London.

[30] Mu¨ ller, K., Mika, S., Ra¨tsch, G., Tsuda, K. and Scho¨lkopf, B. (2001) “An introduction to kernel-based learning algorithms.” *IEEE Trans. Neural Networks* 12 (2):181–202.

[31] Boukharouba, A. and Bennia, A. (2017) “Novel feature extraction technique for the recognition of handwritten digits.” *Applied Computing Informatics* 13(1):19–26.

[32] Jarrett, K., Kavukcuoglu, K., Ranzato, M., and LeCun, Y. (2009) “What is the best multi-stage architecture for object recognition?” In: *Proceedings of IEEE 12th International Conference on Computer Vision (ICCV)* 2146-2153.

[33] Ciresan, D. C., Meier, U., Masci, J., Gambardella, M. L., Schmidhuber, J., Flexible, High-Performance Convolutional Neural Networks for Image Classification, In proceedings of Twenty-Second International Joint Conference on Artificial Intelligence, 1237-1242, 2011.

[34] Ciresan, D., Meier, U. and Schmidhuber, J. (2012) “Multi-column Deep Neural Networks for Image Classification.” *IEEE Conference on Computer Vision and Pattern Recognition* 3642-3649.

[35] Hinton, G. E., Osindero, S. and Teh, Y. W. (2006) “A fast learning algorithm for deep belief nets.” *Neural Computation* 18(7):1527-1554.

[36] Qu, X., Wang, W., Lu, K., and Zhou, J. (2018) “Data augmentation and directional feature maps extraction for in-air handwritten Chinese character recognition based on convolutional neural network.” *Pattern Recognition Letters* 111:9-15.

[37] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2017) “ImageNet classification with deep convolutional neural networks.” *Communications* of the ACM 60(6):84-90.

[38] Zeiler, M.D. and Rob, F. (2014) “Visualizing and understanding convolutional networks.” In: *Proceedings of the European Conference on Computer Vision-2014* 818-833.

[39] Jaderberg, M., Simonyan, K. and Zisserman, A. (2015) “Spatial transformer networks”, In: *Proceedings of the Advances in Neural Information Processing Systems*.

[40] Szarvas, M., Yoshizawa, A., Yamamoto, M. and Ogata, J. (2005) “Pedestrian detection with convolutional neural networks”, In: *Proceedings of the IEEE on Intelligent Vehicles Symposium* 224–229.

[41] Mori, K., Matsugu, M. and Suzuki T. (2005) “Face recognition using SVM fed with intermediate output of CNN for face detection.” in: *Proceedings of the IAPR Conference on Machine Vision Applications*, 410–413.

[42] Özyurt, F., Sert, E., Avci, E. and Dogantekin, E. (2019) “Brain Tumor Detection Based on Convolutional Neural Network with Neutrosophic Expert Maximum Fuzzy Sure Entropy.” *Measurement* in press

[43] Duan, M., Li, K., Yang, C. and Li, K. (2018) “A hybrid deep learning CNN–ELM for age and gender classification.” *Neurocomputing* 275: 448-461.

[44] Wang, X., Sheng, Y., Deng, H. and Zhao, Z. (2019) “CHARCNN-SVM for Chinese Text Datasets Sentiment Classification with Data Augmentation.” *International Journal of Innovative Computing, Information and Control* 15(1):227-246.

[45] Xue, D.X., Zhang, R., Feng, H. and Wang, Y.L. (2016) “CNN-SVM for microvascular morphological type recognition with data augmentation.” *Journal of Medical and Biological Engineering* 36(6):755-764.

[46] Guo, Q., Wang, F., Lei, J., Tu, D. and Li, G. (2016) “Convolutional feature learning and Hybrid CNN-HMM for scene number recognition.” *Neurocomputing*, 184:78-90.

[47] LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998) “Gradient-based learning applied to document recognition.” In: *Proceedings of the IEEE* 86(11):2278-2324.

[48] Ranzato, M., Poultney, C., Chopra, S. and LeCun, Y. (2006) “Efficient learning of sparse representations with an energy-based model.” In: John Platt Bernhard Scholkopf, Thomas Hofmann (Eds.), *Advances in Neural Information Processing Systems*, MIT Press, 1134–1144.

[49] Simard, Y. P., Steinkraus, D. and Platt, C. J., “Best practice for convolutional neural networks applied to visual document analysis.” In: *Proceedings of the International Conference on Document Analysis and Recognition* 2:958–962.