

# A Study to Find Facts Behind Preprocessing on Deep Learning Algorithms

Dr. G. Ranganathan,

Head of the department,

Department of Electronics & Communication Engineering,

Gnanamani College of Technology,

Namakkal, India.

**Abstract:** In the near future, deep learning algorithms will be incorporated in several applications for assisting the human beings. The deep learning algorithms have the tendency to allow a computer to work on its assumption. Most of the deep learning algorithms mimic the human brain's neuron connection to leverage an artificial intelligence to the computer system. This helps to improve the operational speed and accuracy on several critical tasks. This paper projects the blocks, which are required for the incorporation of deep learning based algorithm. Also, the paper attempts to deeply analyze the necessity of the preprocessing step over several deep learning based applications.

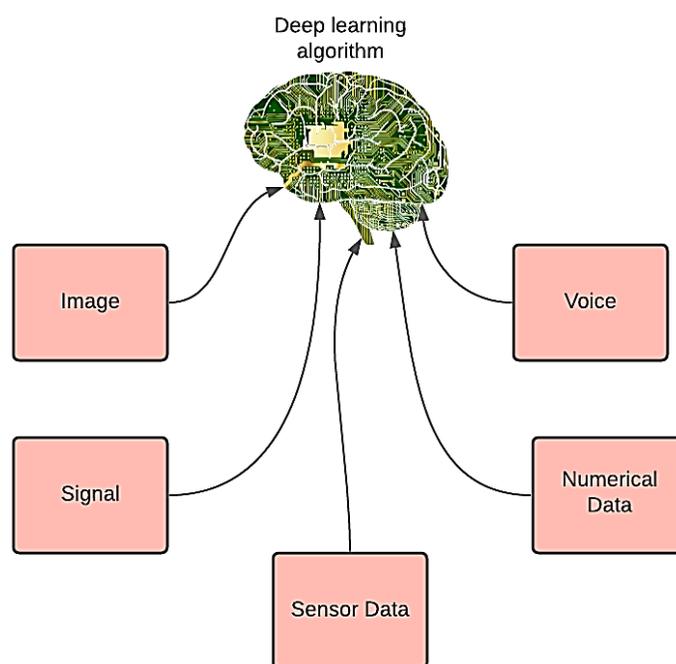
**Keywords:** Preprocessing, image, data, signal, text, classification, EEG, ECG, EMG.

## Introduction

Computers perform some specific task based on the user command. However, due to the improvement of science and technology the researchers were made a computer system to autonomously perform some analytic and arithmetic processes. The applications are getting increased day-by-day due to its compatibility/size reduction of both the hardware and software components required for a computer system. The accuracy and speed of the computer systems are really incomparable to a human speed and accuracy. A computer never gets tired; never do mistake when sleepy and never take leave. These are the reasons for employing an automated system everywhere instead of human beings.



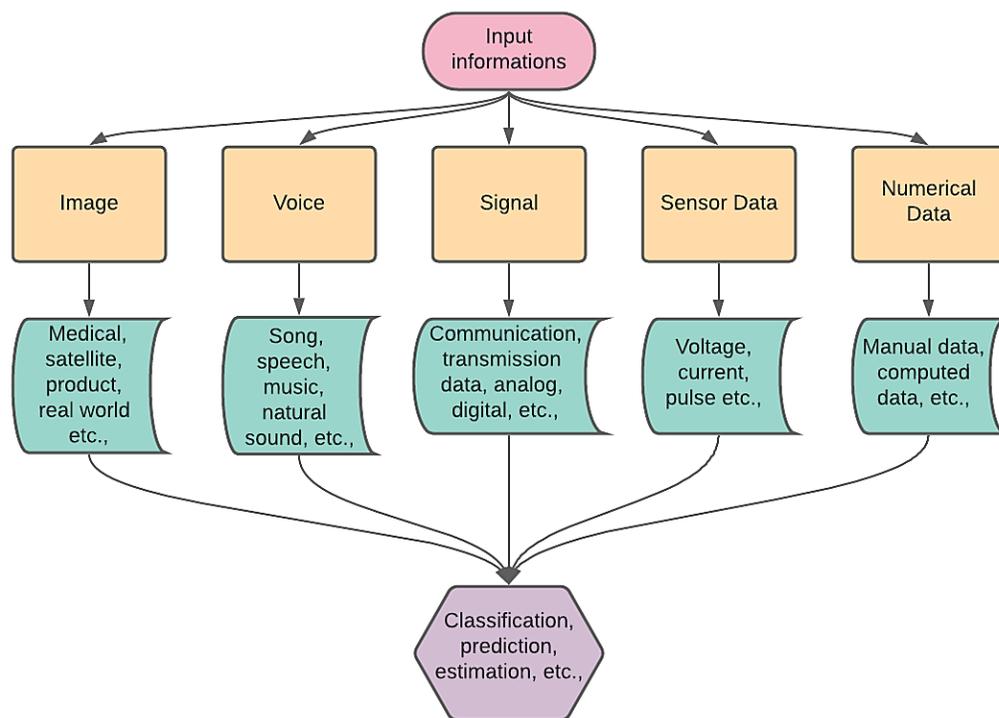
As of now, the computers are made to do the prescribed work. In recent trends, the researchers attempt to give artificial intelligence to a computer machine in order to analyze and work on its own. Many robotic applications are utilizing the artificial intelligence for addressing certain issues. For example, a line following robot is trained to analyze and identify a line, which was drawn in-front of it. Similarly, several medical application robots are trained to visualize the changes occurred over a scanned image and regular image. Figure 1 shows the broad classification of deep learning algorithms input for a reference.



**Figure 1.** Inputs of deep learning algorithm

The deep learning algorithm requires input for analyzing. Same as like of human vision the images are taken into account for several analysis applications. Similarly numerical data, voice, digital signals and sensor outputs are considered as an input for deep learning based algorithms. Figure 2 indicates the application wise classification of input data on deep learning process. However, the inputs are different form the deep learning algorithm, which requires proper preprocessing step for removing the noise . So that, the accuracy of the algorithm gets

increased. Apart from noise removal, the preprocessing step is utilized for marking the area of interest. Therefore, the works are concentrated more on the particular region without any disturbances.

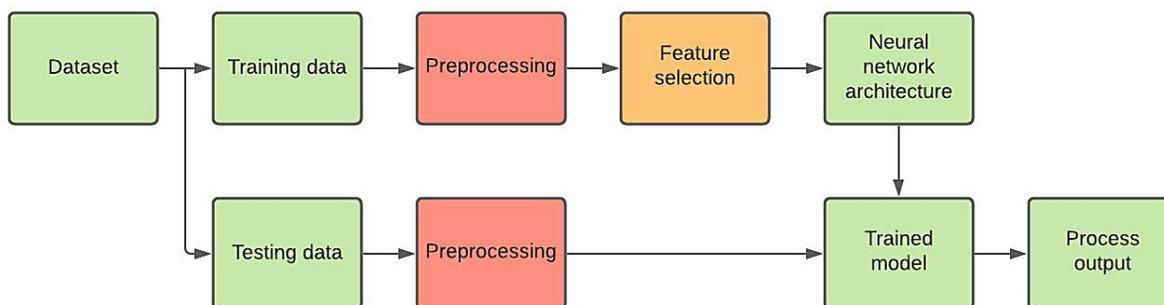


**Figure 2.** Application and outcome of deep learning algorithms

The outcome from deep learning algorithms can be divided into three major categories like classification, prediction and estimation. The classification works are employed to classify two or more categorized images based on the information collected during training process. The prediction kinds of applications on deep learning algorithms that are mostly used for analysis process are obtained from a numerical data. The estimation type of applications is working from the sensor and further the signal data gets transferred to the deep learning architectures.

## Literature Survey

The work flow process of deep learning algorithms on several applications are almost same as like shown in figure 3. Whereas, the dataset splitting and neural network training process will be available for all kind of applications. Also, the feature selection and preprocessing steps are not employed everywhere. The deep learning algorithms can be trained directly with a training data without any preprocessing, when the sample data are clear and noise free. The deep learning algorithms are having the ability to be trained with noisy information. However, it may affect the accuracy on most of the applications. In order to improve the accuracy, the preprocessing steps are widely used for classification and prediction process.



**Figure 3.** Work Flow of Deep Learning Algorithm

Similarly, the feature selection process is also not mandatory for training a deep learning network. However, it is employed in the workflow for concentrating over some important information, which lies over the training samples. Table 1 show the research works, which are employed with a preprocessing step on their workflow. At the same time table 2 indicates the algorithms, which are implemented without any preprocessing work.

**Table 1.** Deep learning based Applications with Preprocessing Techniques

Citation	Area	Approach	Preprocessing	Application	Accuracy
Poloni et al., (2021)	Image classification	SVM	Non-Local Means technique	Alzheimer's disease diagnosis	69.44%
Beeravolu et al., (2021)	Image classification	Deep CNN	Sobel filter	Breast cancer classification	99.06%
Wang et al., (2021)	Image classification	Random Implication Image Classifier	Median filter	Iris disease detection	96.7%
Akhter et al., (2021)	Data classification	Multisize Filters CNN	Stemming	Document classification	95.4%
Lichouri et al., (2021)	Data Classification	BiLSTM	Lemmatization, stemming and POS tagging	Sentiment detection	88.29%
Lichouri et al., (2021)	Data Classification	LSVC	Lemmatization, stemming and POS tagging	Sentiment detection	98.46%
Vijayakumar et al., (2020)	Data classification	Capsule network	Tokenization	Text classification	96.15%
Mitra (2020)	Data classification	SVM, KNN, and NLP	Stemming and tokenization	Sentiment analysis	94.61%
Diker et al., (2020)	Signal classification	Extreme learning machine	Baseline-wandering and 60 Hz noise removal	ECG signal classification	97.5%
Ergün et al., (2020)	Signal classification	KNN	Fusion based preprocessing	EEG classification	88.71%
Heidari et al., (2020)	Image Classification	CNN	Contrast normalization, bilateral and Gaussian low-pass filter	X-Ray classification on COVID 19	94.5%
HaCohen-Kerner et al., (2020)	Data Classification	SMO	Spelling correction, lowercasing, and removal of html tags, punctuation, stop words, & R-repeated characters.	WebKB text classification	95.74%
HaCohen-Kerner et al., (2020)	Data Classification	SMO	Spelling correction, lowercasing, and removal of html tags, punctuation, stopwords, & R-repeated characters.	R8 text classification	95.75%
HaCohen-Kerner et al., (2020)	Data Classification	RF	Spelling correction, lowercasing, and removal of html tags, punctuation, stopwords, & R-repeated characters.	SMS spam classification	98.34%
HaCohen-Kerner et al., (2020)	Data Classification	RF	Spelling correction, lowercasing, and removal of html tags,	Sentiment text classification	78.78%

			punctuation, stopwords, & R-repeated characters.		
Manoharan, S. (2019)	Image classification	SVM	Bayesian Filter	Automobile leak detection	98%
Jacob, I. J. (2019)	Image classification	Capsule networks	Fuzzified image filter	Biometric recognition	99%
Yu et al., (2019)	Signal classification	SVM	Second order Butterworth filter	EMG classification	99.23%
Camps (2018)	Image Classification	Factorial Discriminant Analysis (FDA)	Moving average filtering	Tomato variety classification	94%

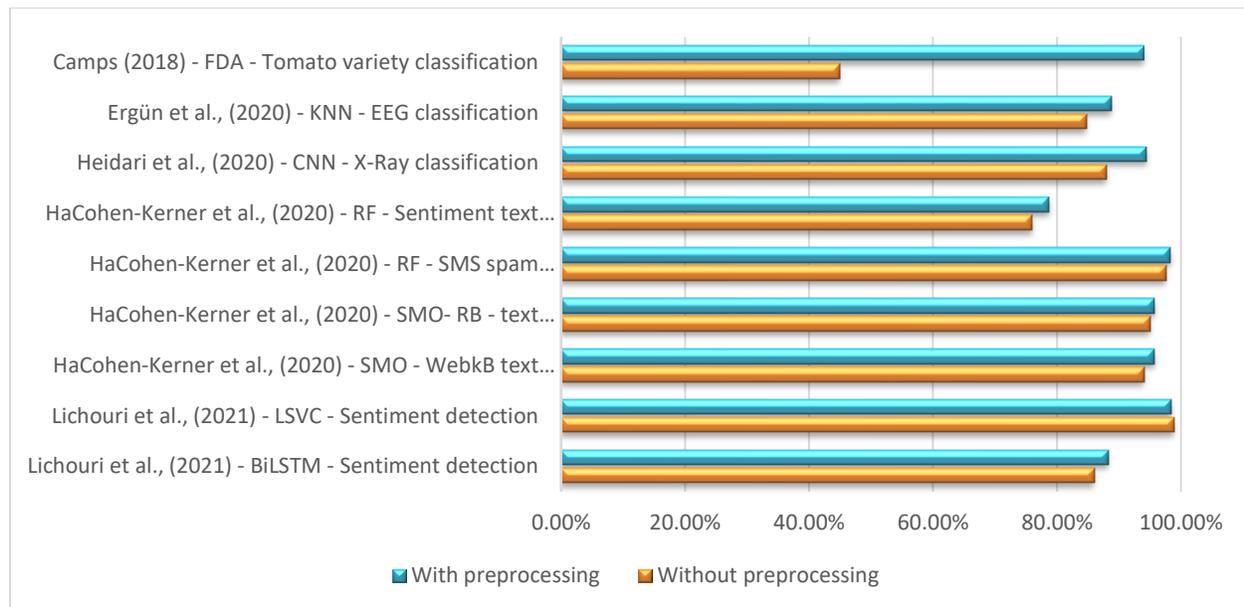
**Table 2.** Deep Learning based Applications without Preprocessing Techniques

Citation	Area	Approach	Application	Attainments
Lichouri et al., (2021)	Data Classification	BiLSTM	Sentiment detection	86.05%
Lichouri et al., (2021)	Data Classification	LSVC	Sentiment detection	98.83%
HaCohen-Kerner et al., (2020)	Data Classification	SMO	WebKB text classification	94.1%
HaCohen-Kerner et al., (2020)	Data Classification	SMO	R8 text classification	94.98%
HaCohen-Kerner et al., (2020)	Data Classification	RF	SMS spam classification	97.62%
HaCohen-Kerner et al., (2020)	Data Classification	RF	Sentiment text classification	75.99%
Heidari et al., (2020)	Image Classification	CNN	X-Ray classification on COVID 19	88%
Ergün et al., (2020)	Signal classification	KNN	EEG classification	84.8%
Rodrigues et al., (2020)	Image Classification	CNN	HEp-2 cell classification	98.28%
Fujita et al., (2019)	Signal classification	CNN	ECG arrhythmia prediction	97.78%
Camps (2018)	Image Classification	Factorial Discriminant Analysis (FDA)	Tomato variety classification	45%

## Discussion

The performances of classification techniques applied to image, data and signal classification techniques are analyzed in table 1 and 2. Based on the accuracy outcome of the classification processes, figure 4 graph has been plotted to observe the performance difference among techniques with preprocessing and without preprocessing. The comparison work that has

been carried out from figure 4 exposes the betterment of the preprocessing technique involved in all kind of applications. A slight performance reduction is observed only on LSVC based sentiment detection process. In remaining approaches, the performances of techniques implemented with preprocessing achieves its betterment.



**Figure 4.** Comparison Analysis of Accuracy on Techniques with Preprocessing and Without Preprocessing

## Conclusion

The deep learning algorithms are having the ability to get trained without utilizing any preprocessing techniques. However, to improve the accuracy efficiency, it requires a preprocessing step. The paper discusses about the importance of preprocessing techniques on deep learning applications. The literature work summarizes the techniques, which are implemented with and without preprocessing step. It projects that, the approaches which are equipped with preprocessing module are comparatively giving better outcome for image, data and signal classification algorithms.

## References

- [1] Akhter, M. P., Jiangbin, Z., Naqvi, I. R., Abdelmajeed, M., Mehmood, A., & Sadiq, M. T. (2020). Document-level text classification using single-layer multisize filters convolutional neural network. *IEEE Access*, 8, 42689-42707.
- [2] Beeravolu, A. R., Azam, S., Jonkman, M., Shanmugam, B., Kannoorpatti, K., & Anwar, A. (2021). Preprocessing of Breast Cancer Images to Create Datasets for Deep-CNN. *IEEE Access*, 9, 33438-33463.
- [3] Camps, C. (2018). Singular approach to penetrometry by preprocessing of digitized force–displacement curves and chemometry: A case study of 12 tomato varieties. *Journal of texture studies*, 49(4), 378-386.
- [4] Chen, Z., Zhou, L. J., Da Li, X., Zhang, J. N., & Huo, W. J. (2020). The Lao text classification method based on KNN. *Procedia Computer Science*, 166, 523-528.
- [5] Diker, A., Avci, E., Tanyildizi, E., & Gedikpinar, M. (2020). A novel ECG signal classification method using DEA-ELM. *Medical hypotheses*, 136, 109515.
- [6] Ergün, E., & Aydemir, O. (2020). A new evolutionary preprocessing approach for classification of mental arithmetic based EEG signals. *Cognitive Neurodynamics*, 14(5), 609-617.
- [7] Fujita, H., & Cimr, D. (2019). Decision support system for arrhythmia prediction using convolutional neural network structure without preprocessing. *Applied Intelligence*, 49(9), 3383-3391.
- [8] HaCohen-Kerner, Y., Miller, D., & Yigal, Y. (2020). The influence of preprocessing on text classification using a bag-of-words representation. *PloS one*, 15(5), e0232525.
- [9] Heidari, M., Mirniaharikandehei, S., Khuzani, A. Z., Danala, G., Qiu, Y., & Zheng, B. (2020). Improving the performance of CNN to predict the likelihood of COVID-19 using chest X-ray images with preprocessing algorithms. *International journal of medical informatics*, 144, 104284.
- [10] Jacob, I. J. (2019). Capsule network based biometric recognition system. *Journal of Artificial Intelligence*, 1(02), 83-94.

- [11] Lichouri, M., Abbas, M., Benaziz, B., Zitouni, A., & Lounnas, K. (2021, April). Preprocessing Solutions for Detection of Sarcasm and Sentiment for Arabic. In Proceedings of the Sixth Arabic Natural Language Processing Workshop (pp. 376-380).
- [12] Manoharan, S. (2019). Image detection classification and recognition for leak detection in automobiles. *Journal of Innovative Image Processing (JIIP)*, 1(02), 61-70.
- [13] Mitra, A. (2020). Sentiment Analysis Using Machine Learning Approaches (Lexicon based on movie review dataset). *Journal of Ubiquitous Computing and Communication Technologies (UCCT)*, 2(03), 145-152.
- [14] Poloni, K. M., de Oliveira, I. A. D., Tam, R., Ferrari, R. J., & Alzheimer's Disease Neuroimaging Initiative. (2021). Brain MR image classification for Alzheimer's disease diagnosis using structural hippocampal asymmetrical attributes from directional 3-D log-Gabor filter responses. *Neurocomputing*, 419, 126-135.
- [15] Rodrigues, L. F., Naldi, M. C., & Mari, J. F. (2020). Comparing convolutional neural networks and preprocessing techniques for HEP-2 cell classification in immunofluorescence images. *Computers in biology and medicine*, 116, 103542.
- [16] Vijayakumar, T., & Vinothkanna, R. (2020). Capsule Network on Font Style Classification. *Journal of Artificial Intelligence*, 2(02), 64-76.
- [17] Wang, Y., & Shan, S. (2021). Accurate disease detection quantification of iris based retinal images using random implication image classifier technique. *Microprocessors and Microsystems*, 80, 103350.
- [18] Yu, M., Li, G., Jiang, D., Jiang, G., Tao, B., & Chen, D. (2019). Hand medical monitoring system based on machine learning and optimal EMG feature set. *Personal and Ubiquitous Computing*, 1-17.

