

A Comprehensive Review on Sentiment Analysis Techniques and Machine Learning Libraries in Image Processing

D.N.V.S.L.S. Indira¹, Ch. Suresh Babu², K. Kranthi Kumar³, Ch. Venkateswara Rao⁴
^{1,2}Associate Professor, ^{3,4}Assistant Professor

^{1,2,3,4}Department of Information Technology, Gudlavalleru Engineering College,
Gudlavalleru, AP, India-521356.

indiragamini@gmail.com¹, sureshdani2004@gmail.com², kk97976@gmail.com³, venkatgecit@gmail.com⁴

Abstract— Big Data is the gateway to social opportunity, dealing with and separating large data action courses (called big data) to discover outlines and other knowledge that accommodates them. Analysis of big data will assist the partnership to better understand the information found within the data and can also help to interpret the data that is most important to business and future business decisions. Sentiment analysis is a method by which the use of natural language processing continues to derive information from the viewpoint of the user. This one remains a task used, for instance, to label individual evaluations as different classifications, positive and negative from the given bit of material. In human decision making, it helps. Machine Learning (ML) has evolved from a piece of Mathematics (Statistics) that only from time to time called computational methodologies, to an autonomous teaching of science that has not only provided the vital basis for learning systems' factual computational standards. This paper is aimed to present the survey of machine learning techniques and sentiment analysis in Distributed Environment and in Image Processing which uses medical data sets.

Keywords : Big Data, Natural Language Processing, Machine Learning, Sentiment Analysis, Image Processing.

I. INTRODUCTION

For a given point, notion examination stays the best approach to remove emotions or assessments from a touch of substance. This causes us to comprehend the perspectives, suspicions, and feelings in the material. In it, the interests of an end client are caught from web content. It includes predicting or inspecting the presentation of mystery information in the substance. This covered information is profoundly useful in placing cooperation into the interests of a client. The motivation behind the last investigation is to decide the perspective of a writer or a speaker for a given subject. Feeling investigation may likewise be applied to sound, photos, and chronicles. Today, the Web has become an indispensable piece of our lives. Web based contributing to a blog objections or long-range casual contact areas are utilized by the more noteworthy bit of the overall population to convey their suspicions about explicit things. Moreover, they utilize these objections to get a handle on what the contemplations of others are. The mining of this information and the extraction of appraisals has subsequently become a fundamental territory of study.

Online users are increasingly expanding, so the gigantic measure of information that is referred to as large information is being generated. The multi dimensional existence of Big Data. There are the following characteristics of big data: variety, velocity, volume, value and veracity. Knowledge is generated in every computerized and web-based existence. Online networks such as Facebook and

Twitter produce a large amount of knowledge about content. Extraordinary esteem is granted to the extraction of useful data and the characterization of tweets.

Calculations for data mining will continue to run in consecutive or parallel conditions. Arrangement calculations that work in a successive domain have problems such as an expansive dataset stockpiling and prolonged execution time. There are a few detriments to the consecutive execution of machine learning calculations for extensive knowledge, such as dependence on information and assignment, non-versatility and a restricted measure of memory. Subsequently, consecutive execution for large datasets does not work adequately. A disseminated situation is used to achieve parallelism to improve the execution of machine learning calculations in contrast to the corresponding situation. There are a few points of interest in a parallel condition, such as information undertaking autonomy, adaptation to non-critical loss, and adaptability. We need huge memory, parallel processing structure, supported libraries, and also synchronization between parallel nodes for faster preparation of vast information.

II. METHODOLOGY FOR SENTIMENT ANALYSIS

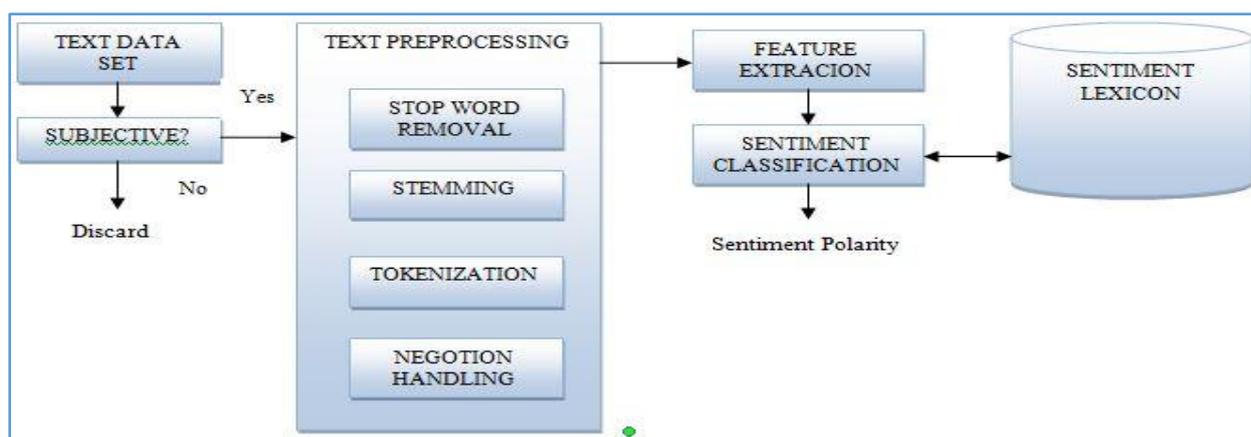


Figure 1: Sentiment Analysis Process

III. IMPORTANT NOTATIONS

A. Subjectivity/Objectivity: -

Most importantly, for performing slant investigation, we need to comprehend the abstract and target text. The abstract content incorporates the feelings where the target proof is put away as the goal text.

B. Polarity: -

The emotional content can be isolated into three classes dependent on sentiments communicated in the content; they are positive, negative and impartial.

C. Sentiment level: -

Notion Analysis should be possible at various levels-

Report Level-In it, the whole record is given a positive, negative or target single limit.

Sentence Level-A sentence level is organized. Each term is independently separated and relegated to the negative, positive or objective. Therefore, the overall file has a few sentences where each sentence has its own furthest point.

Feature Level :-It needs a considerably deeper examination of the content and deals with the defining facts of the expressions or opinions of a sentence and dissects the expressions and orders them as positive, negative or objective. Based on perspective, it is also called an investigation.

Feature Selection : -The key task is to delete appearance from text that is N grams, POS labeling, Stemming, Stop words, Negotiation handling, to perform the classification of sentiment.

D. Sentiment Classification

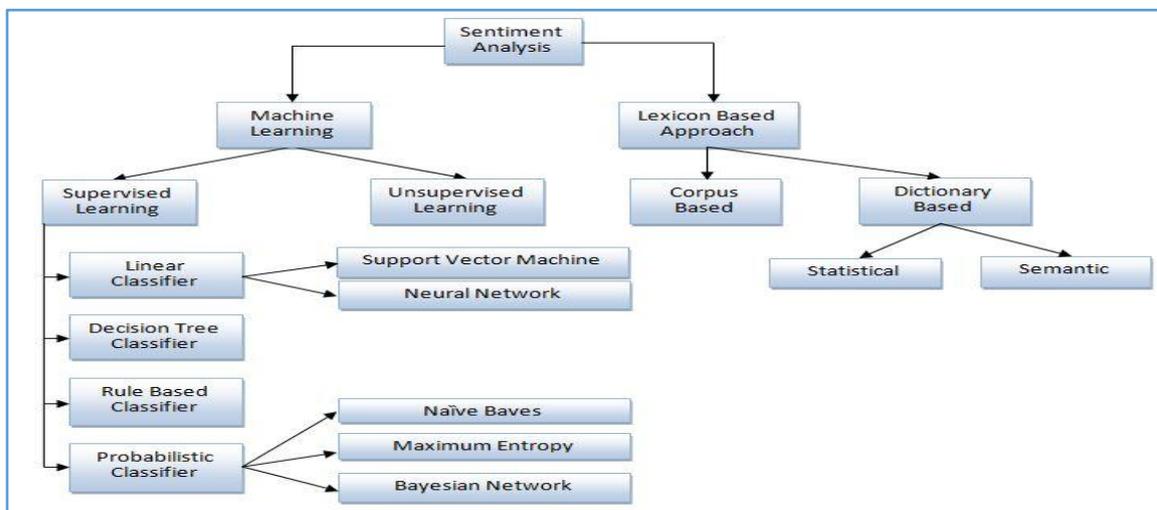


Figure 2: Sentiment Analysis Classification Techniques

In sentiment classification, two methods are commonly used, and those are subjective lexicon and machine learning those are Subjective lexicon, Dictionary based approach

Machine Learning (ML):- ML is a robotized method for grouping; characterization is done utilizing text highlights. The attributes are extricated from the content. There are two sorts of ML, which are regulated learning and solo learning. Utilizing named preparing guides to prepare the gadget. Each class speaks to and has an imprint related with various attributes. Its qualities are coordinated and marked with a class with ideal coordinating when a word shows up.

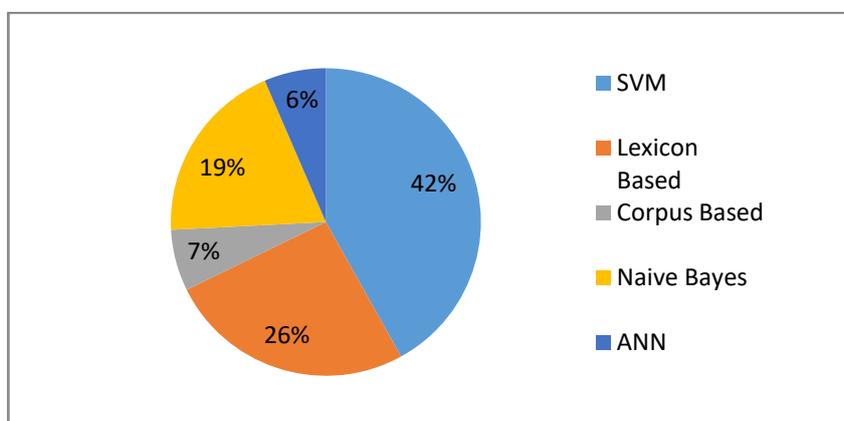


Figure 3- Various Models used in Articles for Text Analytics

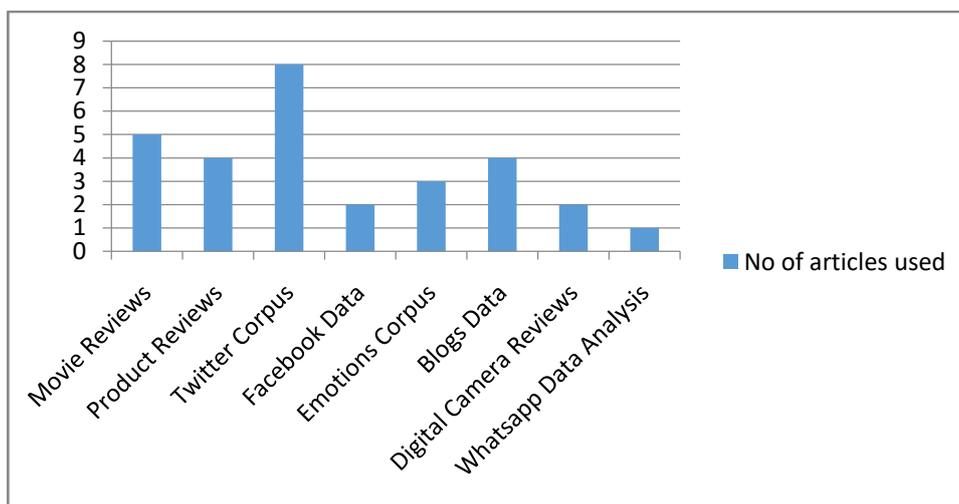


Figure 4- Various datasets used in different articles for text Sentiment Analysis

E. Parallism

Parallel data processing has some critical points of interest over sequential data processing as following:

Speed: it improves the performance and minimises the execution time.

$$\text{Speed (n)} = \text{Time (1)}/\text{Time (n)}$$

Where, Time(1): Execution time of process on single node,

Time(n): Execution time of the process on n nodes.

Scalability:provides scalability to grow Map Reduce and data size of more cluster nodes are available.

$$\text{Scalability (n)} = \text{Time (1,D)}/\text{Time (n,nD)}$$

Where, n is number of cluster nodes,

Time (1, D): Execution time of certain process on D dataset executing on 1 node

Time (n, nD): Execution time of n*D dataset runs on n nodes.

Shared memory: multiple nodes can work on data independently, however, share memory assets so data sharing would turn out to be quick.

F. Approaches to Achieve Parallelism in Machine Learning Algorithm

Three procedures are used to achieve equal execution of AI calculations in circulated climate. These approaches are Data Parallelism, Task Parallelism, and Hybrid Parallelism.

Machine learning algorithm	Data parallelism		Task parallelism	Hybrid parallelism
	Record based	Attribute based		
Naïve bayes classifier	YES	YES	NO	NO
Support vector machine	YES	NO	YES	YES
Decision tree	YES	YES	YES	YES
K-nearest neighbour	YES	YES	NO	NO
K-means	YES	YES	NO	NO
Association rule mining	YES	NO	NO	NO

Table 1: Different parallelism approaches used in machine learning algorithms

G. Machine learning algorithm libraries in Image Processing

ML computation can run in equal or dispersed environment. For running ML computation different Libraries are accessible. Table 2 analyzes different methods that were used in various articles for analyzing medical data set to identify sentiment among images.

S.No.	Year	Methods
1.	2010	DBN,SVM
2.	2010	PCA
3.	2012	Deep CNN
4.	2013	A probabilistic patch-based method, Deep CNN
5.	2014	Deep Belief Networks (DBN) , MRI scans
6.	2014	3-Dimensional CNNs,
7.	2014	
8.	2015	SVM and RF Binary Classifiers
9.	2015	CNNs combined with Support Vector Machine (SVM) and Random Forest (RF)
10.	2016	Stacked Sparse Auto Encoder (SSAE)
11.	2016	CNNs
12.	2016	Used 3 CNNs, each with a different 2-dimensional input patch size, running in parallel to classify and segment MRI brain images
13.	2017	CNN architectures and metrics used in segmentation.
14.	2017	Data-Driven Techniques
15.	2017	The CNNs for the task of visual sentiment prediction
16.	2018	Deep Learning Algorithms (DBN)
17.	2019	Decision Tree
18.	2019	SVM
19.	2019	CNN
20.	2020	Invented speckle-modulating OCT to generate low speckle images to be used as the ground truth based on GAN.
21.	2020	Combined CNN with pigment epithelial detachments and Convolutional Denoising Auto Encoders to do the segmentation of retinal low-cost OCT images.
22.	2020	Used DenseNet201 and a special training method to do the binary classification for retinal disease
23.	2020	Used CNN to segment CNV in OCT angiography.

Table 2: Various Algorithms Used In Different Articles for Identifying Sentiment in Medical Images

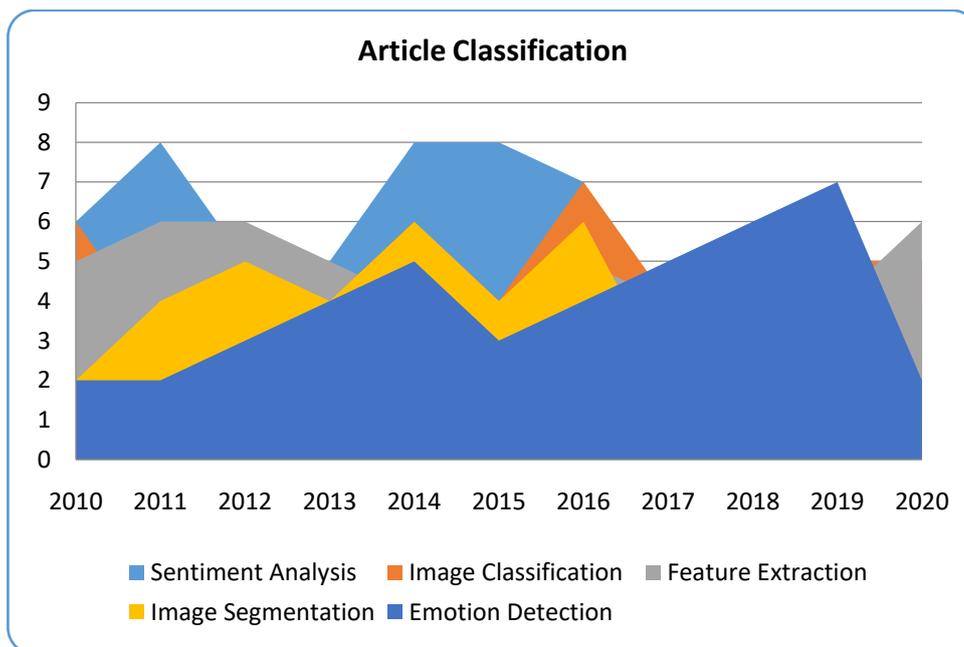


Figure 5: Different Approaches for calculating sentiments in image data sets from 2010 to 2020.

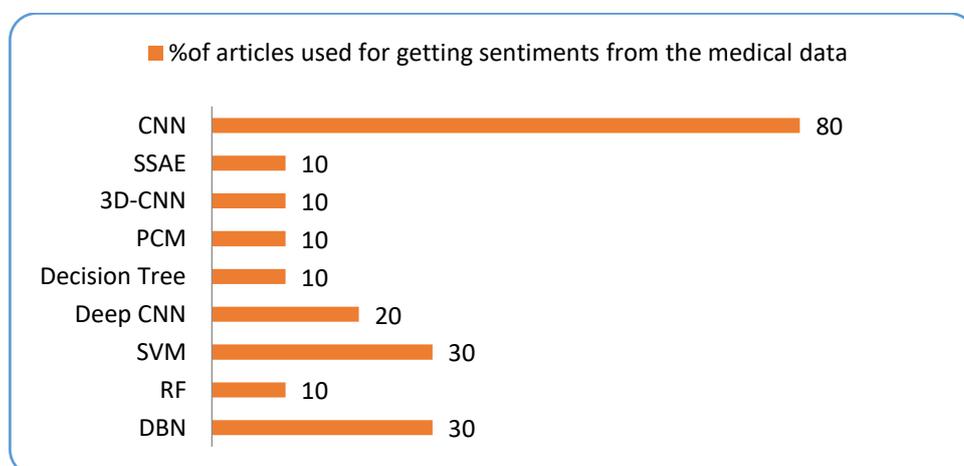


Figure 6: % of articles used for getting sentiments from the medical data

IV. CONCLUSION

This paper presents an overview of sentiment analysis in image data sets. Actually Sentiment Classification techniques are used to mine the text data in olden days. But recent improvements in machine learning algorithms and convolution neural networks, authors are trying to get sentiment from the images also. CNN is the best way to examine images in Python. It provides almost all ML algorithms for processing of images in different ways like feature extraction, image classification, emotion extraction, etc. This overview closes with that feeling characterization is as yet an open field for inspect. SVM and innocent bayes are most well known calculations for slant order. There is a great deal of testing issues like ascertaining conclusion score, distinguishing suitable calculation for the dataset in dispersed climate and reconciliation of all these into

one stage. New calculations are needed to coordinate all the three popular Sentiment Analysis, Machine Learning and Image Processing. Hence more examination is needed in this field.

V. REFERENCES

1. P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio and P.-A. Manzagol, "Stacked denoising auto encoders: Learning useful representations in a deep network with a local denoising criterion", *J. Mach. Learn. Res.*, vol. 11, no. 12, pp. 3371-3408, Dec. 2010.
2. J.Machajdik and A. Hanbury, "Affective image classification using features inspired by psychology and art theory", *Proceedings of the International Conference on Multimedia ser. MM '10*, pp. 83-92, 2010.
3. A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks" in *Advances in Neural Information Processing Systems 25*, Curran Associates, Inc., pp. 1097-1105, 2012.
4. H.-C. Shin, M. R. Orton, D. J. Collins, S. J. Doran and M. O. Leach, "Stacked auto encoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1930-1943, Aug. 2013.
5. S. M. Plis et al., "Deep learning for neuroimaging: A validation study", *Front Neurosci.*, vol. 8, pp. 229, Aug. 2014.
6. R. Li et al., "Deep learning based imaging data completion for improved brain disease diagnosis", *Med. Image Comput. Comput.-Assist. Intervent.*, vol. 17, pp. 305-312, Sep. 2014.
7. Medhat, Walaa, Ahmed Hassan, and HodaKorashy. "Sentiment analysis algorithms and applications: A survey" *Ain Shams Engineering Journal* 5.4 :1093-1113, 2014.
8. F. Ciompi et al., "Bag-of-frequencies: A descriptor of pulmonary nodules in computed tomography images", *IEEE Trans. Med. Imag.*, vol. 34, no. 4, pp. 962-973, Apr. 2015.
9. W. Shen, M. Zhou, F. Yang, C. Yang and J. Tian, "Multi-scale convolutional neural networks for lung nodule classification" in *Information Processing in Medical Imaging*, Cham, Switzerland:Springer, vol. 24, pp. 588-599, Jun. 2015.
10. J. Xu et al., "Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images", *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 119-130, Jan. 2016.
11. K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. J. Snead, I. A. Cree and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images", *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1196-1206, May 2016.
12. P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. de Vries, M. J. N. L. Benders and I. Išgum, "Automatic segmentation of MR brain images with a convolutional neural network", *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1252-1261, May 2016.

13. Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin and B. J. Erickson, "Deep learning for brain MRI segmentation: State of the art and future directions", *J. Digit. Imag.*, vol. 30, no. 4, pp. 449-459, 2017.
14. G. Litjens et al., "A survey on deep learning in medical image analysis", *Med. Image Anal.*, vol. 42, pp. 60-88, Dec. 2017.
15. V. Campos, B. Jou and X. Gir-I-Nieto, "From pixels to sentiment: Fine-tuning CNNs for visual sentiment prediction", *Image Vis. Comput.*, vol. 65, pp. 15-22, 2017.
16. G. Wang, J. C. Ye, K. Mueller and J. A. Fessler, "Image reconstruction is a new frontier of machine learning", *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1289-1296, Jun. 2018.
17. AjitkumarShitole and ManojDevare, "TPR PPV and ROC based Performance Measurement and Optimization of Human Face Recognition of IoT Enabled Physical Location Monitoring", *International Journal of Recent Technology and Engineering*, vol. 8, no. 2, pp. 3582-3590, July 2019, ISSN 2277-3878.
18. Sahar A. El_Rahman, FeddahAlhumaidiAlOtaib and Wejdan Abdullah AlShehri, "Sentiment Analysis of Twitter Data", 2019 International Conference on Computer and Information Sciences (ICIS), 2019, ISBN 978-1-5386-8125-1.
19. A. Kalavani and D. Thenmozhi, "Sentiment analysis using deep learning techniques", *Int. J. Recent Technol. Eng.*, vol. 7, no. 6S5, pp. 1-7, 2019.
20. Z. Dong, G. Liu, G. Ni, J. Jerwick, L. Duan and C. Zhou, "Optical coherence tomography image de-noising using a generative adversarial network with speckle modulation", *Journal of Biophotonics*, 2020
21. T. Kepp, H. Sudkamp, C. von der Burchard, H. Schenke, P. Koch, G. Hüttmann, et al., "Segmentation of Retinal Low-Cost Optical Coherence Tomography Images using Deep Learning", 2020.
22. A. Suzuki and Y. Suzuki, "Deep learning achieves perfect anomaly detection on 108 308 retinal images including unlearned diseases", 2020.
23. J. Wang, T. T. Hormel, L. Gao, P. Zang, Y. Guo, X. Wang, et al., "Automated diagnosis and segmentation of choroidal neovascularization in OCT angiography using deep learning", *Biomedical Optics Express*, vol. 11, no. 2, pp. 927-944, 2020