International Journal of Engineering & Technology, *7 (4.39) (2018) 915-920*

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Analysis and Review of Extraordinary Machine Learning Approaches

Nitin Tyagi[1]*, Sandeep Gupta[2], Arun Pratap Srivastava[3], Shashank Awasthi[4]

[1]Research Scholar, Shri Venketeshwara University, Gajraulla, India
[2]JIMS-EMTC, Greater Noida, India
[3, 4]G L Bajaj Institute of Technology & Management, Greater Noida, India
**Corresponding author E-mail : nitinnitin8218@gmail.com*

Abstract

Machine learning is utilized in the medicinal imaging field, including PC supported finding, picture division, picture enrollment, picture combination, picture guided treatment, picture explanation, furthermore, picture database recovery. Deep learning techniques are arrangements of calculations in machine learning, which endeavor to consequently take in numerous levels of portrayal and deliberation that assistance, comprehend information. This thusly prompts the need of comprehension and looking at the attributes of profound learning approaches, with the end goal to have the capacity to prosecute and process the strategies in legitimate manner. The aim of this article is to assess deep learning (DL) strategies for therapeutic area also to comprehend if DL techniques (random recursive support vector machines (R^2SVM), stacked sparse auto-encoders (SSAE), stacked denoising auto-encoders (SDAE), *K*-means deep learning calculation) beat other best in class approaches (K-closest neighbor, support vector machines, greatly randomized trees) on two arrangement undertakings, where different strategies are assessed for a transcribed digit (MNIST). Results demonstrate that the SSAE, SDAE what's more, the SVM accomplish the most noteworthy exactness among all assessed methodologies on both datasets.

Keywords: *Random Recursive Support Vector Machine (R^2SVM), Stacked Sparse autoencoder (SSAE), Stacked Denoising Autoencoder (SDAE), Convolutional Neural Network (CNN), Support Vector Machine (SVM)*

1. Introduction

Restorative imaging includes all procedures which permit to picture covered up parts of the body for clinical purposes or therapeutic research. A few strategies that have been produced throughout the most recent decades are known for this errand and they generously enhance current medicinal basic leadership the same number of dismal changes of the body occurs inside, undetectable to each exposed human eye. One noticeable precedent is stomach magnet resonance imaging (MRI) in radiology which permits a plainly visible understanding into the stomach some portion of the body. Second regular models are tissue microarrays (TMA) in pathology which permit a tiny amplification of human cells to imagine cell modifications. Such medicinal pictures are generally utilized for recognizable proof, analysis and evaluating of different kinds of ailments, issue or knobs.
Deep learning techniques are an arrangement of calculations in machine realizing, which take in different levels of portrayal and reflection that assistance comprehend information. Larger amount deliberations are characterized starting lower-level ones, further more perplexing capacities can be scholarly. Specifically, in the event that a capacity may be minimally spoken to by a profound design, a similar capacity could require an amazingly substantial design if profundity of this engineering is made extra facile [1]. Moreover, such short review shows that computative writing tenders a great assortment of DL outlooks. This article will play out an exact assessment of delegate outlooks, and talk about the ends from such discoveries.
In ongoing years, fast development of computerized PCs, mixed media, and capacity frameworks has brought about extensive picture and mixed media content vaults. Clinical and indicative investigations are additionally profiting by these advances in computerized capacity and substance preparing. The doctor's facilities having indicative and investigative imaging offices are delivering expansive measure of imaging information along these lines causing a tremendous increment underway of restorative picture accumulations. Consequently, advancement of a compelling medicinal picture recovery framework is desired to help therapists in perusing these extensive datasets. To encourage the procedure of generation and administration of such vast medicinal picture databases, numerous calculations for programmed investigation of therapeutic pictures have been proposed in writing [2-6].
The point of this article is to assess machine training strategies in the medicinal area along with to examine if DL techniques beat best in class approaches. All the more unequivocally, the techniques are assessed on two characterization errands, where one assignment comprises of arranging pictures of a medicinal dataset (figured tomography pictures of the pleura). The rest grouping assignment includes a data subdivision of manually written notations which empowers a chance to distinguish the distinctions of the strategies amidst a regular issue and an errand in any restorative area. Close to an assessment as far as exactness, the strategies are likewise thought about as far as runtime. Here, not just the total estimations of the precision and the runtime are contrasted and one another, yet in addition the impact of the hyper parameter framework and the quantity of utilized precedents after execution are inspected. Moreover, the highlights which were found out by the DL approaches are assessed as well. As an example, it's in-

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

spected if the DL approaches are fruitful in training issue particular highlights consequently. Another target of this proposition is to work out down to earth suggestions for the assessed techniques, if they could benefit enhance the appositeness and rectification of the assessed techniques in eventual. The real commitments of this work are triple,

1. A dataset is precisely gathered that is multimodal and spreads an extensive variety of medicinal imaging target regions.
2. A profound learning system is displayed and prepared on an accumulation of medicinal pictures.
3. The scholarly highlights are utilized to display an exceedingly proficient medicinal picture recovery framework that endeavours for a vast gathering of multimodal data division.

Whatever is left of the article is sorted out as takes after. Related work is exhibited in Section 2, the proposed system is examined in Section 3, exploratory outcomes are appeared and talked about in Section 4 and an end is displayed in Section 5.

2. Literature review

This section focuses conventional machine learning (ML) points, for example, supervised learning (SL) and unsupervised learning (USL). Due to many years of research exercises, is to a great degree huge, just a little extent can be considered in this theory [7]. In any case we cover the themes, which are required keeping in mind the end goal to comprehend the particular ML approaches which were assessed and to more readily evaluate the discoveries in proposed theory. Above all else, with the end goal of a better comprehension of the contrasts between the methodologies, a review of learning composes is given. Accordingly the decision of hyper-parameters is examined, because of the way that they must be decided for all techniques which are assessed inside this work.

2.1. Machine learning

In ML, the regular objective is to discover an aligning from info examples to yield esteem. For example, we have pictures of items as info information (spoken to by pixel intensity) and right names (one for each kind of question) as comparing yield esteems. At that point the point of the calculation is to take in this mapping (from the examples to the yield esteem), to have the capacity to anticipate the right yield of another information test [8].

2.2. Type of algorithms

There are distinctive ML settings, as represented in Figure 1 and are clarified beneath. The attributes of the informational collection and the particular undertaking decide the calculation type, as depicted in the accompanying.

Supervised learning: The SL issue is an exemplary ML the hang of setting. The preparation information is comprised of tuples (x_i; y_i), where x_i is the info and y_i the relating destination vector [9]. The instance where the objective esteem is distinct, for example, the notation acknowledgment issue, in which the pictures are graphed to a limited number of distinct classes, is known as grouping issue. In an event that the objective esteem is ceaseless, the assignment is called relapse issue. A case of a relapse issue could be the expectation of the cost of a home (consistent yield esteem), in which the info factors are quantity of living zone and the rooms.

Unsupervised learning: In USL issues, any preparation information contains precedents of the information vectors with no comparing target esteems. There are unique targets in unsupervised learning issues, for example, grouping, thickness estimation and representation. The objective of bunching is to find gatherings of comparable precedents, based on estimated or saw similitude between the precedents. To decide the conveyance of the information inside the info area is the reason in thickness assessment. As a representation, the information is anticipated bottomward, from an immense-spatial arena, to a few measurements [8].

Semi-supervised learning (SSL) and **Self-taught learning (STL):** These two methodologies are utilizing unlabeled information in supervised learning undertakings and are along these lines somewhere between SL and USL. The pair likelihood to acquire calculations gain from untagged information and a test to get (adequate) marked information for SL errands are inspiring variables for these methodologies. Primary thought is to give the calculation a lot of unlabeled information to take in a decent component portrayal of the info and afterward feed the named information to the calculation to comprehend the administered undertaking based on the scholarly component portrayal. In the semi-supervised learning the hang of setting, the preparation informational collection can be separated into two sections: the information tests $X_l = \{x_1, ..., x_l\}$ with relating marks $Y_l = \{y_1, ..., y_l\}$ and the information tests $X_u = \{x_{l+1}, ..., x_{l+u}\}$ where the marks are not known [11]. Then again, the self-taught picking up setting is the more great setting, since it doesn't accept, conversely to the semi-regulated getting the hang of setting, that the unlabeled information X_u pursues a similar circulation (or then again class names) as the marked information X_l [10].

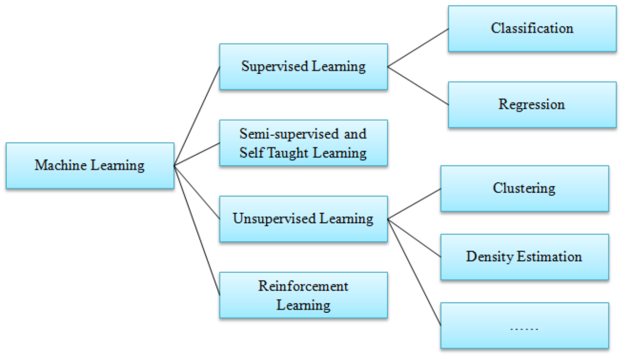

Fig 1: Overview of Machine Learning Setting

To stretch the contrast between the two specified methodologies, we think about a precedent, where the point is to recognize pictures of felines and pictures of mutts. The two methodologies utilize marked pictures, whereby every precedent is either a picture of a canine or a picture of a feline. In the event that there are unlabeled information models that are largely pictures of either a feline or a pooch (however are not marked), at that point this setting would be called semi-administered. Conversely, in self-educated realizing there is an unlabeled informational index, which comprises of arbitrary pictures (maybe downloaded off the Internet) and is in this manner, drawn from an unexpected dissemination in comparison to the marked informational index.

Reinforcement learning: Reinforcement learning is worried about the issue of how to associate with the earth and to pick reasonable activities in a given circumstance, keeping in mind the end goal to augment the reward. Rather than directed learning, there are no models with an ideal yield given however rather, the calculation ought to realize which activities to execute in a particular circumstance by connecting with nature (procedure of experimentation) [8].

2.3. Hyper parameters

Picking *hyper-parameters* is in this manner formally proportionate to show choice, i.e. picking the most fitting quality/calculation in the given arrangement of qualities/calculations [12]. Hyper-parameters can be nonstop (for example training rate) or discrete (for example the quantity of neurons in a single layer) what's more, can be viewed as an outside control catch. To call attention to the distinction between hyper parameters what's more, parameters, we consider a precedent where we might want to prepare a polynomial capacity to fit a particular capacity. The polynomial capacity takes the form:

$$y(x,w) = w_0 + w_1 x + w_2 x^2 + ... + w_M x^M = \sum_{j=0}^{M} w_j x_j \tag{1}$$

here M represents the degree of the polynomial and w_i represents the coefficients of polynomial $(w = w_0, w_1, ..., w_M)$. M must be picked manually before the preparation of the polynomial. Interestingly, the coefficients of polynomial w_i are adjusted within the preparation technique itself. Thus M may be viewed as a hyper-parameter, though the polynomial coefficients are parameters. For all training calculations which are to be analyzed in this article, hyperparameters have to be balanced. At times, for example, neural systems, the quantity of hyper-parameters can be generally high (at least ten), which entangles the issue. In this proposal, the particular hyper-parameters of the different learning calculations are examined as well.

2.4. Model selection

Distinctive machine learning models indicate contrasts regarding unpredictability and the quantity of free parameters, which can be balanced. Adjacent to the model-choice itself, the qualities for the parameters of the model must be resolved. All the more accurately, the fundamental target in this setting is to pick estimations of the complementary criterion as the best prescient execution on recent information is accomplished [13].
To assess the prescient execution of various qualities for the parameters, the preparation set is certainly not a decent pointer, because of the issue of overfitting. Rather, if there is sufficient information, the dataset can be isolated into three sections. The display is prepared with the initial segment of the information and diverse hyper-parameter settings, whereby this part is called preparing set. The second piece of the information isn't utilized for preparing and consequently appropriate to analyze the different hyper-parameter settings and select the one with the best assessed prescient execution. This autonomous informational index is called *approval set*. To at long last assess the chose display regarding speculation, a third piece of the information, called test set, is important. This is on account of the approval set was utilized to choose the last model and is thusly no longer altogether autonomous from the model (overfitting to the approval information can happen).

2.5. Underfitting and overfitting

For better understanding we think about a precedent, where we need to fit the capacity $sin(2\pi x)$ with a polynomial of request M. The request M of the polynomial must be picked, to decide the many-sided quality of this direct model. As appeared in Figure 2, the higher request polynomial (M=9) demonstrates an ideal fit to the preparation information, yet sways fiercely and speaks to the 'genuine' work badly. This is called overfitting [29]. Then again, the lower arrange polynomials ($M = 0, M = 1$) are not sufficiently intricate to adapt the fundamental patterns in preparation information (as appeared in Figure 2) and are in this way rather poor portrayals of the 'genuine' work as well. This is called *underfitting*.

2.6. Role of preprocessing

Data normalization: The basic transforming advance in ML comprises of figuring the average esteem (everything being equal) of a model and insinuate this from each measurement of the precedent. Being pictures, it can be viewed as standardization of the luminosity. In a few errands the by and large brilliance data of a picture is unimportant, wherefore this standardization is fitting (e.g. question acknowledgment assignments [14]).
Another standardization stride is to register the predictable error of a precedent and separation the model by this processed esteem. Pictures have the worth of difference standardization (as specified in Coates et al. [15]).

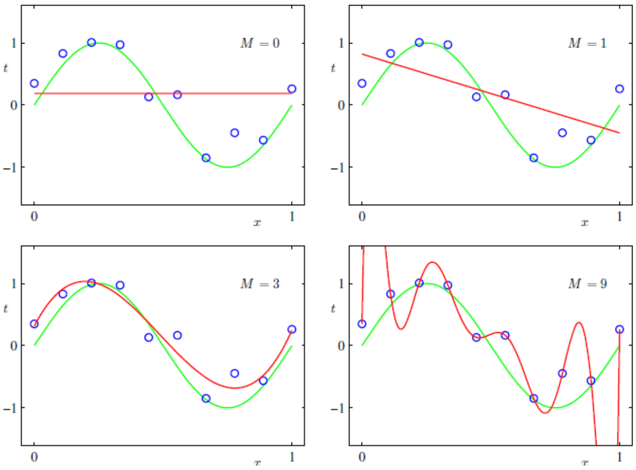

Fig 2: Graphs of polynomials for various degrees of M, shown as red colour. The function $\sin(2\pi x)$ is shown as a green color. The blue points tell about the data points of the learning set ($\sin(2\pi x)$ along with additional stochastic noise)[8].

3. Implementation

This section focuses on the execution of the previously mentioned strategies and proposed work. Fundamentally, MATLAB is utilized as the essential dialect to execute the methodologies which were talked about in the past segments. Adjacent to MATLAB, the product devices and structures LIBSVM, LIBLINEAR, and Tensor flow are utilized to get the coveted usefulness.. Along these lines, the examination of hyper-parameters should be possible unified and the outcomes can be envisioned in a clear way.

3.1. K- nearest neighbour

As MATLAB is utilized as an essential dialect, we chose to utilize the '*knnclassify*' work given in MATLAB as 'Bioinformatics Toolbox' [16]. This capacity gives a chance to decide the quantity of neigh *K*, the separation cadent and administer of how to arrange information point if a no larger part vote occurs. For the examinations of this article, the K-Nearest Neigh approach is assessed utilizing the *Euclidean separation* as the separation cadent. On the off chance that no larger part referendum is accessible, the closest neigh amid the fixing bunches is utilized to rapture the tie. Along these lines, the main criterion which can be shifted in examinations is the quantity of neighs.

3.2. Support vector machine (SVM)

This article utilizes the LIBSVM programming bundle as the execution of SVM. LIBSVM is an athenaeum which depends on the coding dialect C / C++ also is portrayed in detail in [17]. This very well may be utilized for SVM order, SVM relapse and dispersion assessment. It gives an interface to MATLAB, so it very well may be effortlessly coordinated into our cipher.
The primary explanation behind utilizing the Radial Basis Function (RBF) piece is that it has just a single extra hyper-parameter which must be chosen. This implies the many-sided quality of the model determination with the RBF part is easier than with the polynomial portion (in light of the fact that the polynomial bit

contains more hyperparameters). The another logic is that, in contrast with the polynomial bit, RBF piece displays less analytical troubles [18]. Besides, a sigmoid piece isn't legitimate under some parameters [19].

3.3. R^2SVM

The LIBLINEAR bundle is quicker than LIBSVM for preparing straight SVMs. Thinking about a precedent, where five overlay cross approval is utilized to find the best criteria value on a data division having 20,242 cases along with component measurement of 47,236 LIBLINEAR catches just nearly three seconds for preparing the direct SVM, the preparation using LIBSVM utilizes nearly 350 seconds [41]. Besides, it has been guaranteed that the arrangements of the two bundles with regard of exactness are about the equivalent. In the previously cited model, the cross approval exactness is 96.8% using LIBSVM arrangement and 97.0% using LIBLINEAR arrangement. As expressed in [20], the direct SVM allocators are prepared in a one versus all mold, this article chose to prepare the direct classifiers in a one versus all mode, also. As specified earlier, no unmistakable advantage of single technique and the decision relies upon the current issue. Moreover, the LIBLINEAR programming bundle executes the one-versus all order procedure [21], which is the reason the choice to prepare the direct classifiers in a one-versus all mold stays away from a more mind boggling usage and decreases the quantity of potential blunders in the code (on the grounds that no possess code must be composed and tried).

3.4. Extremely randomized tree

For this methodology, the execution utilized in [22], is additionally utilized in this article. In this execution, the default esteems for order, $K = \sqrt{N}$ and $n_{\min} = 2$ are utilized, here N represents the number of characteristics. It is a sensible decision; because the outcomes acquired using the default context of the additional trees calculation demonstrate no huge distinction to the variation, in which the best criterion are found out by cross approval [23]. To put it plainly, the quantity of trees M is the main parameter which is fluctuated.

3.5. Neural network

The usage of the neural system is concluded in the coding dialect Lua, separately using Python. These dialects are utilized, as the ML libraries Tensorflow what's more, it can be utilized inside these programming dialects. The purpose behind executing neural systems with two unique dialects and athenaeum is to have the capacity to confirm the rightness of the executions, as expressed already. On the other hand Pylearn2 can be effectively executed on GPUs; the GPU utilization is considerably more muddled with Tensor flow. Because of it and the way that the execution time of the profound neural system advances is substantially greater in the event that they are not figured on a GPU, we chose to utilize just the Tensor flow.

3.6. K-means deep learning algorithm

As the structure's cipher is composed using MATLAB, this article chose to execute this USL technique utilizing MATLAB too. The perception elements of MATLAB are utilized to confirm the accuracy of the usage. Initially the first fixes and the preprocessed renditions of the chunks are imagined to investigate the preprocessing method. Also, the educated basic cell channels are sketched with the end goal to contrast them and the outcomes introduced in [24]. Furthermore, the gatherings which are produced in the mind boggling cell preparing stage are pictured to have the capacity to contrast them and the gathering in (with the end goal to confirm the rightness of the preparation technique) and to check the parameter-settings (τ, Δ and so forth.). Besides, the usefulness is partitioned into a few free capacities, so the usage is fair and fabricated measured (e.g. the calculation to make a solitary gathering).

4. Experimental setup and result

In this article, a well known and widely used DL tool Tensor flow and MATLAB has been recycled for developing and training the proposed deep learning framework. The simulations have been performed on HP Laptop with Windows 10 having Intel Core i5 CPU with clock speed of 2.40 GHz with a RAM of 6.00 GB. The proposed method has been evaluated in terms of allocation and recovery results.

4.1. Evaluation process

The assessment procedure, or, in other words this work, is appeared in Figure 3. Most importantly the dataset is isolated within three sections: The preparation, the approval and the assessment set. The preparation set is utilized to prepare the techniques with various hyper parameter environments. This article utilizes grid-search to discover the outstanding hyper parameter ambience, as the examination of the outcomes is simpler in contrast with other procedures [12]. The approval set is utilized to look at the hyper parameter environments as far as prescient execution of each technique, in which the model with the outstanding hyper parameter environment (for example the criterion environment with the most noteworthy prescient execution on the approval set) is chosen for each technique. At long last, the speculation exhibitions of the chose models (one for each technique) are assessed utilizing the test set. The models of the named preparing set are likewise utilized as an unsupervised data division (by disregarding the names), as three of the assessed mechanisms require unsupervised pictures. This procedure of miniature choice and ascertaining the speculation execution is additionally talked about in Section II.

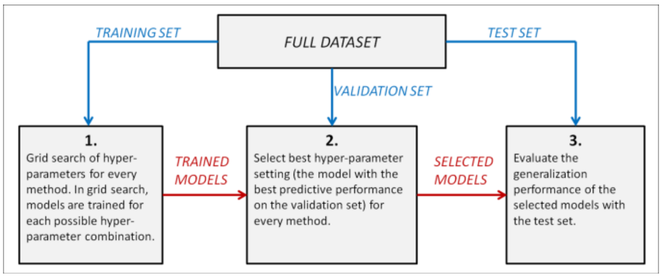

Fig 3: Outline of the assessment procedure utilized in proposed work. On the other hand the "demonstrate stream" is indicated with red, the "information stream" is featured with blue.

4.2. MNIST data division

The Modified NIST (MNIST) data division is a directory of 28-by-28 picture elements transcribed notation pictures also, is a sub division of the NIST directory [25]. The MNIST preparing division contains 30,000 pictures from NIST's Special Directory 1 (SD1), 30,000 pictures from the NIST's Special Directory 3 (SD3) and the examination division has a extent of 10,000 pictures (5,000 for every directory). As expressed, it is important to construct the new data division MNIST out of the NIST's directory, on the grounds that the aftereffects of tests ought to be autonomous of the part of the total data division into preparing and examination division, with the end goal to have the capacity to reach sensible determinations. This contention is conceivable, since initially SD-3 was assigned as a preparation division, SD1 as a examination division, and SD3 is made out of precedents which are less demanding to perceive than that ones in SD1 [50].
The MNIST preparing division contains pictures from around 250 scholars, on the other hand the pictures of the MNIST examination

division start from different journalists. The first dual-matched (high contrast) pictures from the NIST directory were standardized to fit within a 20-by-20 picture elements, in which the viewpoint proportion was kept for all models amid the indicated resizing procedure. As an enemy of associating strategy was utilized inside this standardization method, the subsequent pictures additionally contain dim levels. The estimate standardized precedents were focused in the last 28-by-28 picture by registering the inside of weights of the picture elements (for 20-by-20 picture) and deciphering the picture as the focal point of weight is situated at the focal point of the last 28-by-28 picture [25]. Arbitrarily chose pictures using the MNIST data division are appeared in Figure 4.

(a) (b)

Fig 4: Haphazardly chose models using the MNIST directory of manually written notations. (a) Examples from the preparation division. (b) Examination division examples.

In this area, the exactnesses of the strategies, acquired on the MNIST data division, are introduced furthermore, analyzed. An outline is presented in Table 1, in which the examination correctness (exactnesses registered on the examination division) of all assessed strategies are appeared. The pattern of 10% signifies the uncovered exactness of a basic miniature which arranges all precedents arbitrarily (in the examination division). As the examination division of the MNIST directory displays 10 ranks, the benchmark is 10%. This standard surveys the nature of the assessed models. Exactness is an approach to quantify and analyze the speculation execution of the strategies. It is ascertained as the quantity of right anticipated precedents partitioned by the aggregate number of models.

While the SDAE shows the outstanding examination exactness with 98.55%, the SAE approach has the third most astounding examination exactness of 98.07%. Then again, SVM approach displays the second most astounding examination exactness of about 98.29%. K-implies DL calculation doesn't enhance the execution to a great degree randomized tree strategy, as the examination precision diminishes from 90.31% to 80.03%.

The perplexity grid of the stacked inadequate auto-encoder approach is represented in Figure 4.4, where the general structure of this disarray grid is illustrative of the perplexity lattices everything being equal. This perplexity lattice demonstrates that there are no classes of the MNIST dataset which are particularly hard to recognize or recognize from each other.

The impact of the quantity of marked models on the examination exactness is shown in Figure 5, where every strategy is featured in its very endemic shading. The relationship is the equivalent for all methodologies: a greater number of named preparing precedents expanded the test exactness. Besides, no bend in Figure 4.20 demonstrates a reasonable union. This demonstrates a higher number of patches would enhance the exactness.

Table 1: An outline of the got examination exactnesses; the chosen frameworks (the ones with the most noteworthy approval precision) were connected on the MNIST examination division. Adjacent to the precision, the table outlines the quantity of changed criterion and the quantity of trajectories (for example the quantity of prepared frameworks) for every strategy.

Method	Test Accuracy (MNIST)	# Varied Parameters	# Settings
Stacked Denoising Auto-Encoder	98.55%	6	216
SVM	98.29%	2	25
Stacked Sparse Auto-Encoder	98.07%	6	324
K-Nearest Neighbor	96.92%	1	3
R^2SVM	93.36%	2	15
Extremely Randomized Trees	90.31%	1	5
K-means Deep Learning Algorithm	80.03%	5	48
Baseline	10.00%	-	-

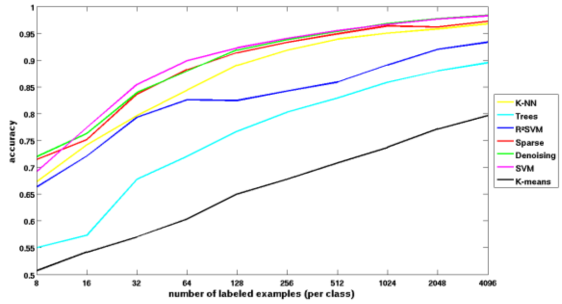

Fig 5: This realistic delineates the connection between's the test precision and the number of marked precedents (which are utilized to prepare the model) for all techniques.

5. Conclusion

Concerning the general execution of the techniques, it tends to be expressed that the SSAE, the SDAE and the SVM accomplished the most elevated exactness amid all assessed methodologies on both data divisions. In the meantime, these three techniques display the most elevated preparing time amid all assessed methodologies on both data divisions. Accordingly these techniques are best if the accessible computational assets permit to utilize them. Interestingly, the K-closest neighbor way and the R^2SVMs show the most limited preparing time on both data divisions, however accomplish a indigent precision than the previously mentioned methodologies. Since the outcomes recommend that the K-closest neigh calculation neglects to explain more confused order assignments, the R^2SVMs are preferable in tending to restorative grouping undertakings if the computational cost is the most vital factor.

Other than these discoveries concerning precision and execution time, the outcomes propose that the K-implies DL calculation, the SSAE and the SDAE are for the most part effective in learning issue particular highlights.

The most forthright eventual development is the association of the two proposed segmentation methods in one single structure. Using hand-operated segmentations as disclosure probabilities, the concurrent segmentation and registration (CSR) framework could be used to construct pleura on the advantageous region of the knob. The use of deformable registration would increase the rigor as well as allow studies on the impact of the knobs on the lung organizations and functional areas. This definable area could then be incorporated as a spatial prior in the CSR method, while cluster assignment could be based on initial detections in the hierarchical structure.

References

[1] Yoshua Bengio, "Learning deep architectures for AI", *Foundations and trends in Machine Learning*, 2(1),1–127, 2009.

[2] M. Mizotin, J. Benois-Pineau, M. Allard, and G. Catheline, "Feature-based brain MRI retrieval for Alzheimer disease diagnosis," in 2012 *19th IEEE International Conference on Image Processing*, 2012, pp. 1241-1244.

[3] G. W. Jiji and P. S. J. D. Raj, "Content-based image retrieval in dermatology using intelligent technique," *IET Image Processing*, vol. 9, no. 4, pp. 306-317, 2015.

[4] M. Ponciano-Silva et al., "Does a CBIR system really impact decisions of physicians in a clinical environment?," in *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, 2013, pp. 41-46.

[5] G. Quellec, M. Lamard, G. Cazuguel, B. Cochener, and C. Roux, "Fast wavelet-based image characterization for highly adaptive image retrieval," *IEEE Transactions on Image Processing,* vol. 21, no. 4, pp. 1613-1623, 2012.

[6] M. M. Rahman, S. K. Antani, and G. R. Thoma, "A learning-based similarity fusion and filtering approach for biomedical image retrieval using SVM classification and relevance feedback," *IEEE Transactions on Information Technology in Biomedicine,* vol. 15, no. 4, pp. 640-646, 2011.

[7] Tom Michael Mitchell, "The discipline of machine learning", *Carnegie Mellon University, School of Computer Science, Machine Learning Department,* 2006.

[8] Christopher M Bishop et al., "Pattern Recognition and Machine Learning", volume 1. *Springer New York*, 2006.

[9] Adam Coates. Demystifying Unsupervised Feature Learning. *PhD thesis, Standford University*, 2012.

[10] Rajat Raina, Alexis Battle, Honglak Lee, Benjamin Packer, and Andrew Y Ng., "Self-taught learning: transfer learning from unlabeled data", In *Proceedings of the 24th international conference on Machine learning*, pp 759–766. ACM, 2007.

[11] Olivier Chapelle, Bernhard Schölkopf, Alexander Zien, et al. "Semi-supervised learning", volume 2, *MIT press Cambridge*, 2006.

[12] Yoshua Bengio, "Practical recommendations for gradient-based training of deep architectures", *In Neural Networks: Tricks of the Trade*, pages 437–478. Springer, 2012.

[13] Pedro Domingos, "A few useful things to know about machine learning", *Communications of the ACM,* 55(10):78–87, 2012.

[14] Nicolas Pinto, David D Cox, and James J DiCarlo, "Why is real-world visual object recognition hard?", *PLoS computational biology*, 4(1):e27, 2008.

[15] Adam Coates and Andrew Y Ng, "Learning feature representations with k-means", *In Neural Networks: Tricks of the Trade*, pages 561–580. Springer, 2012.

[16] MATLAB. Version 8.0.0 (R2012b). *The MathWorks Inc*., Natick, Massachusetts, United States, 2012.

[17] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: "A library for support vector machines", *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[18] Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, et al., "A practical guide to support vector classification", *Technical report*, 2010.

[19] Vladimir Vapnik, "The nature of statistical learning theory" *Springer*, 2000.

[20] Oriol Vinyals, Yangqing Jia, Li Deng, and Trevor Darrell., "Learning with recursive perceptual representations", *In Advances in Neural Information Processing Systems*, volume 4, pages 2834–2842, 2012.

[21] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. LIBLINEAR: "A library for large linear classification", *Journal of Machine Learning Research*, 9:1871–1874, 2008.

[22] René Donner, Bjoern H Menze, Horst Bischof, and Georg Langs, "Global localization of 3d anatomical structures by pre-filtered hough forests and discrete optimization", *Medical image analysis*, 17(8):1304–1314, 2013.

[23] Pierre Geurts, Damien Ernst, and Louis Wehenkel, "Extremely randomized trees", *Machine learning*, 63(1):3–42, 2006.

[24] Adam Coates, Andrej Karpathy, and Andrew Y Ng, "Emergence of object-selective features in unsupervised feature learning", *In NIPS*, volume 25, pages 2690–2698, 2012.

[25] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner, "Gradient-based learning applied to document recognition", *Proceedings of the IEEE*, 86(11):2278–2324, 1998.