IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Volume-6, Issue-3 E-ISSN: 2347-2693

Evolution of Machine Learning Methods for Memography Classification

R. Swathi[1*], R. Seshadri[2]

[1*]CSE, Name of College Sri Venkateswara University,Tirupati,India
[2] CSE, Name of College Sri Venkateswara University,Tirupati,India

[*]*Corresponding Author: swathi.mani08@gmail.com*

Available online at: www.ijcseonline.org

Received: 18/Feb//2018, Revised: 21/Feb2018, Accepted: 23/Mar/2018, Published: 30/Mar/2018

***Abstract*—** In Healthcare and Biomedical sectors, the data is growing more and more, analysing of such medical data accurately will benefits disease detection and early diagnosis. Mammography is the process toward utilizing low-energy X-rays to look at the human cancer for diagnosis and screening. The objective of mammography is the early detection of breast cancer , ordinarily through recognition of trademark masses or macrocalcifications. *Low positive predictive model of mammogram will lead to more no unnecessary* biopsies with benign outcomes. *The accuracy and reliability of prediction mechanisms is important* to reduce the number of biopsies. In this paper, we look at different machine learning algorithms with a specific end goal to predict the performance accuracy. By comparing different algorithms, it has been concluded that deep learning algorithm and Revisiting SVM have highest prediction accuracy among other algorithms studied. Experimental results show this prediction approach is more effective.

***Keywords*—** Deep learning, Machine Learning, Revisiting SVM, SVM.

I. INTRODUCTION

Machine learning is a sub field of Artificial Intelligence. The objective of machine learning by and large is to comprehend the structure of data and fit that data into models that can be comprehended and used by individuals.

Although machine learning is a field in computer science, it contrasts from customary computational methodologies. In conventional computing, algorithms are sets of unequivocally modified directions utilized by computers to compute or issue fathom. Machine learning algorithms rather consider computers to train on information sources of data and utilize measurable examination keeping in mind the end goal to yield esteems that fall inside a particular range. Along these lines, machine learning encourages computers in building models from test data with a specific end goal to mechanize decision making process in light of data inputs.

Any innovation user today has profited from machine learning. Facial recognition innovation enables online networking stages to enable clients to tag and offer photographs of companions. Optical character acknowledgment (OCR) innovation changes over images of content into movable kind. Recommendation engines, controlled by machine learning, recommend what movies or televisions shows to watch next in light of client inclinations. Self-driving cars that depend on machine learning figuring out how to explore may soon be accessible to buyers.

Machine learning is a persistently developing field. Because of this, there are a few contemplations to remember as you work with machine learning systems or analyse the effect of machine learning processes.

At the point when recorded mammographic data is utilized for prediction of accuracy of model, the essential inspiration driving this endeavour is that the knowledge stored in the historical datasets can be utilized to create predictive models. So, prediction approaches must be able to predict model for mammographic data set with high accuracy

Keeping in mind the end goal to create expectation-based programming systems, it is required a high-exact forecast work by either statistical methods such as linear regression and correlation analysis or machine learning methods such as ANN (Artificial Neural Network) and SVM (Support Vector Machine) [1], to predict the effort or duration of projects [2]. Along these lines, expectation approaches must to have the capacity to foresee venture exertion and length, with high exactness. prediction based methodologies require an expectation work that as indicated by the past information of the project, will anticipate the future task exertion and term.

Notwithstanding, because of the huge number of machine learning algorithms.

In this paper, we will think about the execution of a several distinctive Machine learning algorithms to predict value of breast biopsy from mammogram interpretation.

Rest of the paper is organized as follows, Section I contains the introduction of Machine learning and Deep learning, Section II contain the related work of different machine learning algorithms, Section III contain the comparison of machine learning algorithms for classification of mammography, Section IV contain the results of different algorithms on mammography dataset, section V concludes research work with future directions.

II. RELATED WORK

In this section, we will briefly describe different supervised and unsupervised machine learning algorithms and the main characteristics of the machine learning algorithms compared.

A. *Neural Networks*

An Artificial Neuron Network (ANN), prevalently known as Neural Network is a computational model in view of the structure and elements of natural neural systems. It resembles a fake human nervous system for getting, handling, and transmitting data regarding Computer Science.

A neural system learns by changing its parameters. The parameters are the estimations of bias and weights in its neurons. Some neural systems learn continually amid their application, though the clear majority of them have two periods: a training period and an application period. Amid the training period frame a system forms arranged sources of info and alters its parameters. With a specific end goal to enhance its execution, it is guided by some learning algorithm. Once the execution is acceptably exact, or exact, the preparation time frame is done. The parameters of the system are then settled to the educated esteems, and the system begins its time of use for the expected undertaking [3-4].

B. *Decision Trees*

A Decision Tree is a tree (and a sort of coordinated, non- cyclic diagram) in which the nodes speak to choices (a square box), random transitions or terminal nodes, and the edges or branches are double (yes/no, genuine/false) speaking to conceivable ways starting with one node then onto the next. The sort of decision tree utilized for machine learning contains no random transitions. To utilize a decision tree for classification or regression, one snatches a line of information or an arrangement of highlights and begins at the root, and afterward through each ensuing choice hub to the terminal hub. The procedure is extremely instinctive and simple to decipher, which permits prepared decision trees to be utilized for variable determination or more for the most part, highlight designing. To outline this, assume you needed to purchase another auto to drive up an irregular earth street into some arbitrary backwoods. You have a data set of various autos with three highlights: Car Drive Type (Categorical), Displacement (Numeric) and Clearance (Numeric). A case of a scholarly decision tree is below:

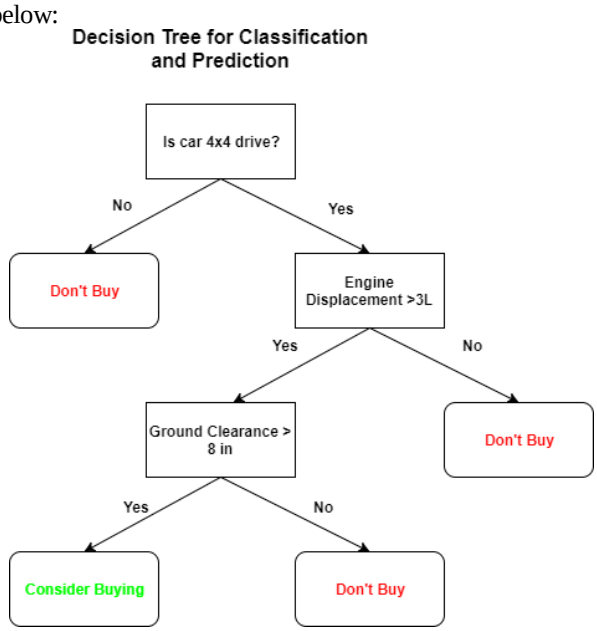

Fig.1

The root or highest node of the tree (and there is just a single root) is the decision node that parts the data set utilizing a variable or highlight that outcomes in the best part metric assessed for every subset or class in the dataset that outcomes from the split. The decision tree learns by recursively splitting the dataset from the root onwards as indicated by the splitting metric at every decision node. The terminal nodes are achieved when the part metric is at a worldwide extremum. Prevalent splitting measurements incorporate the limiting the Gini Impurity (utilized via CART) or expanding the Information Gain (utilized by ID3, C4.5).

C. *Support Vector Machines*

The SVM algorithm is implemented in practice using a kernel. The learning of the hyperplane in linear SVM is finished by changing the issue utilizing some linear algebra. A capable understanding is that the linear SVM can be rethought utilizing the inward result of any two given perceptions, instead of the perceptions themselves. The internal item between two vectors is the whole of the increase of each combine of information esteems.
The equation for predicting a new input by using the dot product between the input (x) and each support vector (xi) is calculated as below:
f(x) = B0 + sum (ai * (x,xi))

This equation that involves in calculating the inner products of a new input vector (x) with all support vectors in training data. The coefficients B0 and ai can be

calculated from the training data by the learning algorithm.

D. Logistic Regression

Notwithstanding the name "Logistic Regression" this isn't a calculation for relapse issues (It is used to predict a real valued output).

Logistic Regression is somewhat like Linear Regression as in both have the objective of assessing the qualities for the parameters/coefficients, so the toward the finish of the preparation of the machine learning model we got a capacity that best portray the connection between the known input and output values. Not at all like Liner Regression, the expectation for the yield is changed utilizing a non-linear function called the logistic function.

Like linear regression, logistic regression improves when we evacuate attributes that are inconsequential to the yield variable and in addition traits that are fundamentally the same as (associated) to each other. The reason Logistic Regression is generally utilized despite the reality of the cutting-edge calculations, for example, profound neural systems this is on the grounds that strategic relapse is extremely effective and does not require excessively computational assets which makes it moderate to keep running on generation [5].

E. KNN

KNN classifier is to group unlabelled perceptions by allocating them to the class of the most similar labelled examples. Attributes of perceptions are gathered for both training and test dataset. For ex, natural product, vegetable and grain can be recognized by their crunchiness and sweetness. To display them on a two-measurement plot, just two qualities are utilized. There can be any number of indicators, and the case can be reached out to consolidate any number of attributes. As a rule, organic products are sweeter than vegetables. Grains are neither crunchy nor sweet. Our work is to figure out which class does the sweet potato have a place with. In this case we pick four closest sorts of nourishment, they are apple, green bean, lettuce, and corn. Since the vegetable wins the most votes, sweet potato is appointed to the class of vegetable. we can see that the key idea of KNN is straightforward.

There are two important concepts in the above example. One is the method to calculate the distance between sweet potato and other kinds of food. By default, the knn() function employs Euclidean distance which can be calculated with the following equation[6,7].

where p and q are subjects to be compared with n characteristics. There are also other methods to calculate distance such as Manhattan distance (8-9).

Another concept is the parameter k which decides how many neighbours will be chosen for kNN algorithm. The appropriate choice of k has significant impact on the diagnostic performance of kNN algorithm. A large k reduces the impact of variance caused by random error but runs the risk of ignoring small but important pattern. The key to choose an appropriate k value is to strike a balance between overfitting and underfitting [10]. Some authors suggest setting k equal to the square root of the number of observations in the training dataset.

F. Navie Bayes

The Bayesian Classification speaks to an supervised learning technique and a measurable strategy for classification. Expect a hidden probabilistic model and it enables us to catch vulnerability about the model principled by deciding probabilities of the results. It can tackle analytic and prescient issues. This Classification is named after Thomas Bayes (1702-1761), who proposed the Bayes theorem. Bayesian classification gives functional learning algorithms and earlier information and observed information can be consolidated. Bayesian Classification gives a helpful point of view to comprehension and assessing numerous learning calculations. It ascertains express probabilities for speculation and it is vigorous to commotion in input information [11].

III. COMPARSION OF MACHINE LEARNING ALGORITHMS FOR CLASSIFICATION OF MAMMOGRAPHY

The main aim of this paper is to compare the effectiveness of different machine learning to predict the accuracy in mammography process. This data contains 961 instances of masses detected in mammograms, and contains the following attributes:

1. BI-RADS assessment: 1 to 5 (ordinal)
2. Age: patient's age in years (integer)
3. Shape: mass shape: round=1 oval=2 lobular=3 irregular=4 (nominal)
4. Margin: mass margin: circumscribed=1 microlobulated=2 obscured=3 ill-defined=4 spiculated=5 (nominal)
5. Density: mass density high=1 iso=2 low=3 fat-containing=4 (ordinal)
6. Severity: benign=0 or malignant=1 (binominal)

Applying several different supervised machine learning techniques to this data set and getting which one yields the highest accuracy as measured with K-Fold cross validation (K=10).

BI-RADS is an appraisal of how sure the seriousness grouping is; it isn't a "predictive" characteristic thus we will dispose of it. The age, shape, margin, and density

attributes are the features that we will build our model with, and "severity" is the classification we will endeavour to foresee in view of those attributes. Although "shape" and "margin" are nominal data types, which sklearn commonly doesn't manage well, they are sufficiently close to ordinal that we shouldn't simply dispose of them. The "shape" for instance is requested progressively from round to irregular. A part of pointless anguish and surgery emerges from false positives emerging from mammogram comes about [12].

IV. EXPERIMENTS RESULTS

If we can construct a better way to interpret them through supervised machine learning, it could improve a lot of lives. Before applying any machine learning algorithm, the data to be cleaned. many rows contain missing data, and there may be erroneous data identifiable as outliers as well. Some techniques such as SVM also require the input data to be normalized first. Many techniques also have "hyperparameters" that need to be tuned. Once we identify a promising approach, see if we can make it even better by tuning its hyperparameters.

V. CONCLUSION

In this paper, we used distinctive machine learning algorithms to build mammography to predict early detection of cancer. At that point we thoroughly thought about various machine learning algorithms. Our exploratory outcomes demonstrate that Neural networks and Revisiting SVM Process gets the better general execution among the contemplated algorithms.

REFERENCES

[1]. K. Nigam, A. Mccallum and T. Mitchell, "Text classification from labelled and unlabelled documents using EM", Machine Learning, vol. 39, no. 2.

[2]. W. Zhang, Y. Yang and Q. Wang, "Handling missing data in software effort prediction with naive Bayes and EM algorithm", PROMISE '11, Banff, Canada, September 20-21,.

[3]. C. L. Martín, A. Chavoya and M. E. M. Campaña, "Use of a Feed Forward Neural Network for Predicting the Development Duration of Software Projects", 2013 12th International Conference on Machine Learning and Applications.

[4]. S. K. Shevade, S. S. Keerthi, C. Bhattacharyya and K. R. K. Murthy, "Improvements to the SMO Algorithm for SVM Regression", IEEE Transactions ON Neural Networks, vol. 11, no. 5, September.

[5]. Short RD, Fukunaga K. The optimal distance measure for nearest neighbour classification. IEEE Transactions on Information Theory 1981; 27:622-7. 10.1109/TIT.1981.1056403

[6]. Weinberger KQ, Saul LK. Distance metric learning for large margin nearest neighbour classification. The Journal of Machine Learning Research 2009; 10:207- 44.

[7]. Cost S, Salzberg S. A weighted nearest neighbour algorithm for learning with symbolic features. Machine Learning 1993; 10:57-78. 10.1007/BF00993481

[8]. Breiman L. Random forests. Machine Learning. 2001; 45:5-32. 10.1023/A:1010933404324

[9]. Zhang Z. Too much covariates in a multivariable model may cause the problem of overfitting. J Thorac Dis 2014;6: E196-7.

[10]. Lantz B. Machine learning with R. 2nd ed. Birmingham: Packt Publishing; 2015:1.

[11]. J. A. Lopez, J. L. Berral, R. Gavalda and J. Torres, "Adaptive on-line software aging prediction based on machine learning", in Procs. 40th IEEE/IFIP Intl. Conf. on Dependable Systems and Networks, June 28, July 1, pp. 497-506.

[12]. Y. Wang, and I. H. Witten, "Inducing Model Trees for Continuous Classes", the European Conf. on Machine Learning Poster Papers.

Authors Profile

Mrs. R Swathi did Bachelor of Technology from University of JNTU, Hyderabad in 2006 and Master of Technology from JNTUA University in year 2011. Currently pursuing Ph.D. in Sri Venkateswara university and working as Assistant Professor in Department of Computer Science and Engineering at SV college of Engineering, Tirupati. She has published many research papers in reputed international journals and conferences including IEEE and it's also available online. Her main research work focuses on Big Data Analytics and Machine Learning. She has 10 years of teaching experience.

Dr. R Seshadri did Bachelor of Technology from Nagarjuna University, Guntur in 1981 and ME from PSG college, Coimbatore in year 1984. Awarded Ph.D. from Sri Venkateswara university in the year of 1998 and working as Director cum professor in Department of Computer Science and Engineering at Sri Venkateswara Engineering college, Tirupati. He has published many research papers in reputed international journals and conferences including IEEE and it's also available online. His main research work focuses on Computer networks and its security, Adhoc networks, Image processing, Big Data Analytics and Machine Learning. He has 30 years of teaching experience.

