ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

Stock Market Prediction System using Back Propagation andC4.5 Algorithm

Kalyan Dani[1], Dr.Basant Tiwari [2]

Research Scholar, Department of CS, TIT, Bhopal (M.P.), India [1]

Professor & Head, Department of CS, TIT, Bhopal (M.P.), India [2]

ABSTRACT: Stock market prediction is essential and of great interest because successful prediction of stock prices may promise smart benefits. These tasks are highly complicated and very difficult. Many researchers have made valiant attempts in data mining to devise an efficient system for stock market movement analysis. In this paper, we have developed an efficient approach to stock market prediction by employing fuzzy C-means clustering and artificial neural network. This research has been encouraged by the need of predicting the stock market to facilitate the investors about buy and hold strategy and to make profit. Firstly, the original stock market data are converted into interpreted historical (financial) data i.e. via technical indicator.

Based on these technical indicators, datasets that are required for analysis are created. Subsequently, fuzzy-clustering technique is used to generate different training subsets. Subsequently, based on different training subsets, different ANN models are trained to formulate different base models. Finally, a meta-learner, fuzzy system module, is employed to predict the stock price. The results for the stock market prediction are validated through evaluation metrics, namely mean absolute deviation, mean square error, root mean square error, mean absolute percentage error used to estimate the forecasting accuracy in the stock market. Comparative analysis is carried out for single Neural Network (NN) and existing technique neural. The obtained results show that the proposed approach produces better results than the other techniques in terms of accuracy.

KEYWORDS:Stock market prediction, Rate of Change (ROC), Money Flow Index (MFI), Relative Strength Index (RSI), stochRSI, ultimate oscillator, MSE RMSE, MAPE, MAD

I. INTRODUCTION

A stock market is a public market for companies or for people to raise money. Stock market helps companies to buy or sell their shares. The price of shares depends upon the demand and supplies of shares. This process of buying and selling of shares is called trading, only the Listed Companies can carry out trading.

Stock market prediction is the process of trying to determine the future stock value of a company. The successful prediction of a stock's future price could yield significant profit. Stock price movements are governed by the theories random walk hypothesis and efficient market hypothesis.

The Forecasters of stock market focus on developing approaches which successfully fore-cast/predict stock prices using well defined trading strategies. A successful prediction model is the one which works with best accuracy having minimum input requirements and least complex model. Investors and government organizations rely on forecasting tools to guard against risks and to monitor market fluctuations. For researchers, these serve as a reference for studies of financial issues like pricing of financial derivatives and portfolio selection.

Generally, the stock traders nowadays depend on Intelligent Trading Systems which help them in predicting prices based on various situations and conditions, which in turn help them in making instantaneous investment decisions. Stock market values are very dynamic and susceptible to quick changes.

From researchers and investors, financial prediction and trading presents a demanding task that attracts great interest because success may result in substantial re-wards. On the other hand, since the financial market is a highly complex and dynamic system, predicting the financial market is not an easy task which involves the actual actions taken by the millions of investors and institutions. By the predictability of the financial market, many investors are persuaded and they try to make profit through exploiting the analysis of financial data [1, 26]. Predicting current

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

concern of factories and manufacturing companies is the desire of investors, auditors, financial analysts, governmental officials, employees and managers. More prominently, within stock market research, it is believed that the information from periodical reports and annual reports can manipulate the price of a stock, especially for unexpected earnings or unexpected loss surprises. The complexity of the task has raised questions on whether the stock market price can be predicted. However, stock market investors believe stock prices can be expected, and pro t can be made through exploiting assorted technical or fundamental analysis, in addition to momentum strategies (buy when market is bullish, sell when market is bearish).

II. RELATED WORK

The price variation of stock market is a very dynamic system that has intrigued analysis from many disciplines. Two common analytical approaches are fundamental analysis and technical analysis. A fundamental analysis relies on the statistics of the macroeconomics data such as interest rates, money supply, in stationary rates, and foreign exchange rates as well as the basic financial status of the company. After taking all these factors into account, the analyst will then decide of selling or buying a stock. A technical analysis is based on the historical financial time-series data. However, financial time series exhibit quite complicated patterns (for example, trends, abrupt changes, and volatility clustering) and such series are often nonstationary, whereby a variable has no clear tendency to move to an axed value or a linear trend.

Stock price prediction has always been a subject of interest for most investors and Professional analysts. Nevertheless, finding out the best time to buy or to sell has remained a very di cult task because there are too many factors that may influence stock prices. Many papers have dealt with input selection when it comes to mapping financial indexes and stocks. Inputs have been divided into two distinct types of inputs, financial and political (which tend to be qualitative) and these inputs have been discussed in earlier researchers. Quah and Srinivasan (1999) identified 5 key factors that will affect the stock price movement namely yield, liquidity, risk, growth, and momentum. Izumi and Ueda (1999) mentioned that macroeconomic factors such as inaction and short-term interest rate have direct impacts on the stock returns.

As stock market data are highly time-variant and are usually in a nonlinear pat-tern, predicting the future trend (i.e., rise, decrease, or remain steady) of a stock is a challenging problem. Analysis and prediction of the stock market behavior have been accompanied by predictions of the behavior of the prices. To put forward future behavior some of the approaches rely on charts of the prices, volumes, and visual human analysis of these diagrammatic representations. Others manipulate the historical values of the time series to calculate technical indicators. The value, or values, of one or more of these are used to suggest good times for buying or selling stock. Both Chartist techniques and the use of indicators are technical models that use only information gained through the trading history of a stock. On the contrary, a basic model looks at the past financial performance of a company, the behavior of the economy, and the industry to which a company belongs. In predicting the future performance, some also use knowledge of the past performance of the directors. Other models mix both technical and fundamental aspects [10, 25]. The indecisive nature of the stock market requires the use of data mining techniques like clustering for stock market analysis and prediction.

Uncovering market trends, planning investment strategies, identifying the best time to purchase the stocks and what stocks to purchase are included in stock market. For approaching these enormously complex and dynamic problems with data mining tools, financial institutions produce huge data sets that build a foundation [3, 4]. To predict stock market along with the development of artificial intelligence, especially machine learning and data mining, ever more researchers try to build automatic decision-making systems. Among these approaches, soft com-putting techniques such as fuzzy logic, neural networks, and probabilistic reasoning (which includes genetic algorithms, chaos theory, etc.) draw most consideration due to their abilities to handle ambiguity and noise in stock market [12, 6]. In both trend analysis and forecasting stock data mining has given encouraging results by using these certain artificial intelligence techniques [21, 27]. In addition, to accomplish better prediction results for forecasting incorporating emerging Artificial Intelligence techniques such as neural networks and/or fuzzy logic with the data mining methods are of wide interest.

Feng Qiu et al , discussed both theoretical and real-life experiments to evaluate the approach. In the theoretical experiments, the user interests estimated by learning algorithm can be used to accurately predict her view on the importance of web pages, which is expressed by her Personalized Page Rank, showing that method is effective and

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

easily applicable to real life search engines. This method performed between 25%33% better than Topic-Sensitive. Page Rank, which turned out to be much better than Page Rank. In the future they plan Prediction of stock market price is done with the help of fuzzy C-means clustering and artificial neural networkevaluation matrices parameters such as mean absolute deviation, mean square error, root mean square error and mean absolute percentage error are used to estimate the forecasting accuracy in the stock market prediction.

III. PROPOSED SYSTEM

The main objective proposed system is to reduce the risk to an individual buyer or seller on each transaction. Primary market deals with the current issues of securities and also securities are brought the share directly from the companies. But the secondary market, securities are bought and sold the shares among investors. Secondary market deals with excellent securities. This market is made of organized exchanges and it has trading of stocks, where orders are transmitted for exchange. All the trading of stocks is maintained and Guided by the exchanges. The rules and regulations are set down by the exchanges.

Back propagation Algorithm

The objective of the training is to minimize the divergence between real data and the output of the network. This principle is referred to as ***Supervised Learning.***

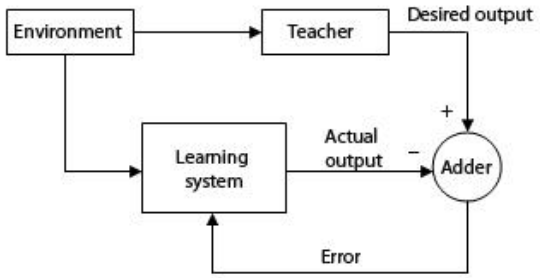

Fig: 4.1 simplified Back propagation algorithm model.

In a step by step manner the error guides the network in the direction towards the target data. The back propagation algorithm belongs to this class and can be described as "an efficient way to calculate the partial derivatives of the network error function with respect to the weights". ***Fig.*** shows the general model for back propagation algorithm.

Most people would consider the Back Propagation network to be the quintessential Neural Net. Actually, Back Propagation 1, 2,3 is the training or learning algorithm rather than the network itself. The network used is generally of the simple type. These are called Feed-Forward Networks or occasionally Multi-Layer Perceptrons (MLPs).

- First apply the inputs to the network and work out the output remember this initial output could be anything, as the initial weights were random numbers.
- Next work out the error for neuron B. The error is What you want What you actually get
- Change the weight. Let W AB be the new (trained) weight and WAB be the initial weight.

Calculate the Errors for the hidden layer neurons. Unlike the output layer we can't calculate these directly (because we dont have a Target), so we Back Propagate them from the output layer (hence the name of the algorithm). This is done by taking the Errors from the output neurons and running them back through the weights to get the hidden layer errors. For example if neuron A is connected as shown to B and C then we take the errors from B and C to generate an error for A.

Having obtained the Error for the hidden layer neurons now proceed as in stage 3 to change the hidden layer weights. By repeating this method we can train a network of any number of layers.

The back propagation algorithm falls into the general category of gradient descent algorithms, which intend to find the minima/maxima of a function by iteratively moving in the direction of the negative of the slope of the function

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

to be minimized/maximized. The main objective is to minimize the error function. The average error function to be minimized (error density).

The main steps using the Back propagation algorithm as follows:

Step 1: Feed the normalized input data sample, compute the corresponding output;
Step 2: Compute the error between the output(s) and the actual target(s);
Step 3: The connection weights and membership functions are adjusted;
Step 4: IF Error > Tolerance THEN go to Step 1 ELSE stop.

C4.5 Algorithm

The algorithm constructs a decision tree starting from a training set T S, which is a set of cases, or tuples in the database terminology. Each case specifies values for a collection of attributes and for a class. Each attribute may have either discrete or continuous values. Moreover, the special value unknown is allowed, to denote unspecified values. The class may have only discrete values. We denote with C1 To CN Class the values of the class.

The C4.5 algorithm constructs the decision tree with a divide and conquer strategy. In C4.5, each node in a tree is associated with a set of cases. Also, cases are assigned weights to take into account unknown attribute values. At the beginning, only the root is present, with associated the whole training set T S and with all case weights equal to 1:0. At each node the following di- vide and conquer algorithm is executed, trying to exploit the locally best choice, with no backtracking allowed.

In doing classification with C4.5, the concepts of entropy and correlation coefficient need to be explained in brief. Entropy is a measure of uncertainty among random variables in a collection of data or in other words entropy provides information about the behavior of random processes used in data analysis. Correlation coefficient has its uses as a chief statistical tool in data analysis finding the relationship between variable sets. Different ways of calculations have been introduced to boost the efficiency of the correlation coefficient among which are Kendall, Pearson's and Spearman's correlation coefficients. There are several test options with WEKA providing data classification such as training set, supplied test set, percentage split and cross validation.

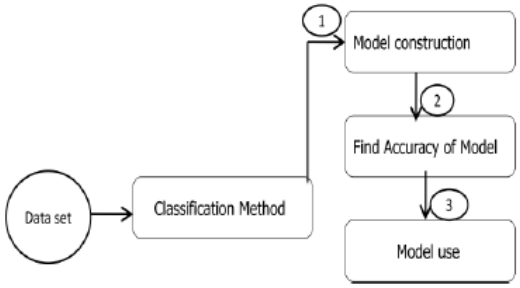

Fig: Classification Steps

III. SIMULATION RESULTS

Results are analysed on the basis of different performance metrics. Graphs shown below shows simulation results are according to network and pause time model i.e. varying number of entry and changing time respectively.

Comparison of MSE, RMSE, MAD and MAPE:

Figure 1 shows comparison of mean square error, Root Mean square error, Mean absolute deviation and Mean Absolute percentage Error. So the comparison shows when time goes high all MSE,RMSE,MAD and MAPE performance goes high.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

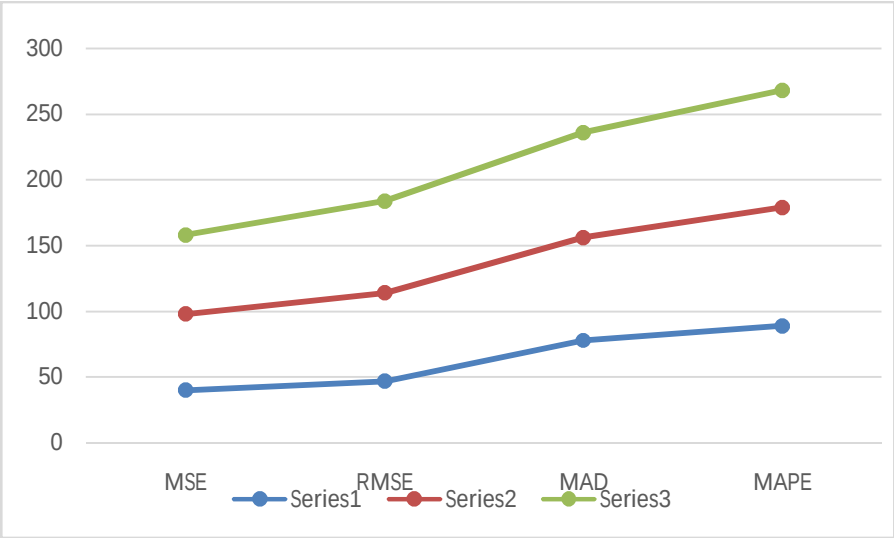

Figure 6.1 Comparison of MSE, RMSE, MAD and MAPE

6.2.2 Comparison of fuzzy model and Back propagation model:

Figure 6.2 shows comparison of Fuzzy Model and back propagation model. It seems that fuzzy model gives low performance than back propagation model that's the reason for prediction stock market value use back propagation model

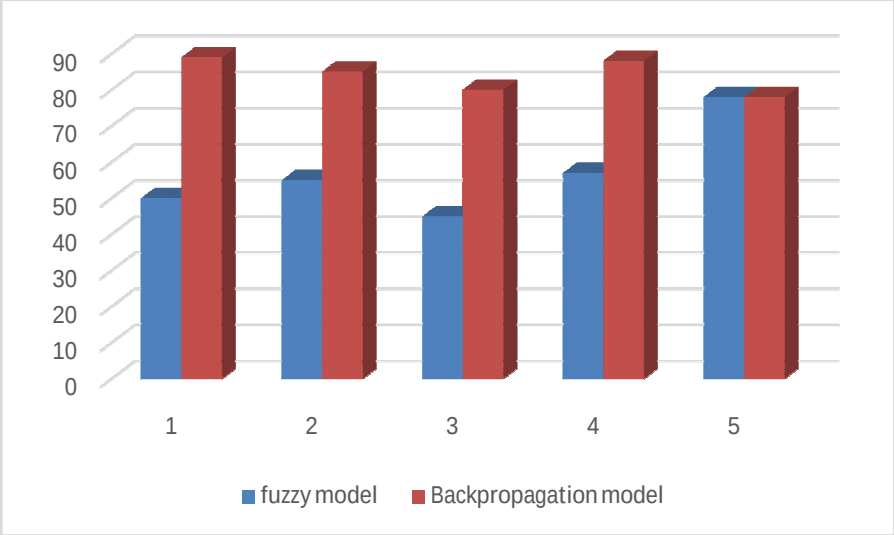

Figure. 6.2 Comparison of fuzzy model and Back propagation model

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

Comparison of ID3 and c4.5 algorithm

Figure 3 shows comparison of ID3 and C4.5 algorithm. C4.5 algorithm gives higher performance and overcome the problem of ID3

The classifier is trained and tested first. Then the resulting decision tree or rule set is used to classify unseen data. C4.5 is the newer version of ID3. C4.5 algorithm has many features like:

- Speed - C4.5 is significantly faster than ID3 (it is faster in several orders of magnitude)
- Memory - C4.5 is more memory efficient than ID3
- Size of decision Trees – C4.5 gets smaller decision trees.
- Rule set - C4.5 can give rule set as an output for complex decision tree.
- Missing values – C4.5 algorithm can respond on missing values by _infinity
- Over fitting problem - C4.5 solves over fitting problem through

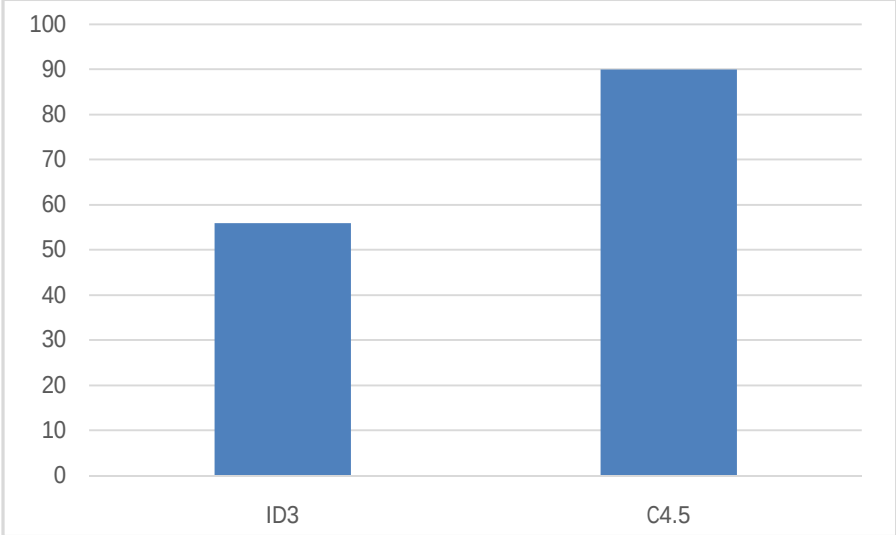

Figure 3 Comparison of ID3 and c4.5 algorithm

IV. CONCLUSION

Prediction of stock market is significant technical index as a consequent to predict the stock price. The results for the stock market prediction is validated through evaluation metrics, namely, mean absolute deviation, mean square error, root mean square error, mean absolute percentage error used to estimate the forecasting accuracy in the stock market. The comparative analysis is carried out on single Neural Network (NN) and existing technique neural. The obtained results show that the proposed approach produces better results than the other techniques in terms of accuracy.

As future work, focusing on adding more features that are specific for each concept, This research is just a beginning and the long term goal is to predict the trend of the price variation by including various influential factors such as macro-economic change, political reasons, fundamental analysis and the technical index . . ., etc. As a result, the system can be further applied for the daily trading purpose.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

REFERENCES

[1] Fei Wang, Noah Lee, Jianying Hu, Jimeng Sun, ShahramEbadollahi, And Andrew F. Laine,‖A Framework For Mining Signatures From Event Sequences And Its Applications In Healthcare Data‖, *IEEE Transactions OnPattern Analysis And Machine Intelligence*, Vol. 35, No. 2, February 2013.

[2] B. Cao, D. Shen, J.T. Sun, X. Wang, Q. Yang, and Z. Chen, ―Detect and Track Latent Factors with Online Nonnegative Matrix Factorization,‖ *Proc. 20th Int'l Joint Conf. Artificial Intelligence*, pp. 2689-2694, 2007.

[3] F.R.K. Chung, *Spectral Graph Theory. Am. Math. Soc*., 1997.

[4] C. Ding, T. Li, and M.I. Jordan, ―Convex and Semi-Nonnegative Matrix Factorizations,‖ *IEEE Trans. PatternAnalysis and Machine Intelligence*, vol. 32, no. 1, pp. 45-55, Jan. 2010.

[5] M. Dong, ―A Tutorial on Nonlinear Time-Series Data Mining in Engineering Asset Health and Reliability rediction: Concepts, Models, and Algorithms,‖*Math. Problems in Eng*., vol. 2010, pp. 1- 23, 2010.

[6] J. Eggert and E. Korner, ―Sparse Coding and NMF,‖ *Proc. IEEE Int'l Joint Conf. Neural Networks*, vol. 2, pp. 2529-2533, 2004.

[7] W. Fei, L. Ping, and K. Christian, ―Online Nonnegative Matrix Factorization for Document Clustering,‖ *Proc. 11thSIAM Int'l Conf. Data Mining*, 2011.

[8] C. Fevotte and J. Idier, *Algorithms for Nonnegative Matrix Factorization with the Beta-Divergence, arXiv*:1010. 1763, 2010.

[9] P.O. Hoyer, ―Non-Negative Matrix Factorization with Sparseness Constraints,‖ *J. Machine Learning Research*, vol. 5, pp. 1457-1469, 2004.

[10] P.O. Hoyer, ―Non-Negative Sparse Coding,‖ *Proc. 12th IEEE Workshop Neural Networks for Signal Processing*, 2002.

[11] Y.R. Ramesh Kumar and P.A. Govardhan, ―Stock Market Predictions—Integrating User Perception for Extracting Better Prediction a Framework,‖*Int'l J. Eng. Science, vol. 2, no. 7*, pp. 3305-3310, 2010.

[12] D.D. Lee and H.S. Seung, ―Learning the Parts of Objects by Non-Negative Matrix Factorization,‖ *Nature, vol*. 401, no. 6755, pp. 788-91, 1999.

[13] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, ―A Symbolic Representation of Time Series, with Implications for Streaming Algorithms,‖ *Proc. Eighth ACM SIGMOD Workshop Research Issues in Data Mining and KnowledgeDiscovery*, pp. 2-11, 2003.

[14] J. Mairal, F. Bach Inria Willow Project-Team, and G. Sapiro, ―Online Learning for Matrix Factorization and Sparse Coding,‖ *J. Machine Learning Research*, vol. 11, pp. 19-60, 2010.

[15] F. Moerchen, ―Time Series Knowledge Mining Fabian,‖ PhD thesis, 2006.

[16] F. Moerchen and D. Fradkin, ―Robust Mining of Time Intervals with Semi-Interval Partial Order Patterns,‖*Proc.SIAM Conf. Data Mining*, pp. 315-326, 2010.

[17] F. Mo¨rchen and A. Ultsch, ―Efficient Mining of Understandable Patterns from Multivariate Interval Time Series,‖*Data Mining and Knowledge Discovery*, vol. 15, no. 2, pp. 181- 215, 2007.

[18] P. OGrady and B. Pearlmutter, ―Discovering Convolutive Speech Phones Using Sparseness and Non-Negativity,‖ *Proc. Seventh Int'l Conf. Independent Component Analysis and Signal Separation*, pp. 520- 527, 2007.

[19] R. Andrew Russell, ―Mobile Robot Learning by Self-Observation,‖ *Autonomous Robots*, vol. 16, no. 1, pp. 81-93, (2004).

[20] J. Shlens, G.D. Field, J.L. Gauthier, M. Greschner, A. Sher, A.M. Litke, and E.J. Chichilnisky, ―The Structure of Large-Scale Synchronized Firing in Primate Retina,‖*J. Neuroscience: The Official J. Soc. for Neuroscience*, vol. 29, no. 15, pp. 5022-5031, 2009.