Lung Cancer Detection using Artificial Neural Network

M. Jerin Jose*, B.R. Janani**, K. Janani Priya***, J. Jeevitha**** & S. Swathi*****

*Assistant Professor, Department of Electronics and Communication Engineering, S.K.P Engineering College, Tamil Nadu, INDIA.
E-Mail: jerin108jose{at}gmail{dot}com
**UG Scholar, Department of Electronics and Communication Engineering, S.K.P Engineering College, Tamil Nadu, INDIA.
E-Mail: Jananibaskar37{at}gmail{dot}com
***UG Scholar, Department of Electronics and Communication Engineering, S.K.P Engineering College, Tamil Nadu, INDIA.
E-Mail: jananiprya25{at}gmail{dot}com
****UG Scholar, Department of Electronics and Communication Engineering, S.K.P Engineering College, Tamil Nadu, INDIA.
E-Mail: dharshinijeevitha{at}gmail{dot}com
*****UG Scholar, Department of Electronics and Communication Engineering, S.K.P Engineering College, Tamil Nadu, INDIA.
E-Mail: swathi.s0496{at}gmail{dot}com

Abstract**—**Computed tomography technology is one of the important technology techniques, which shows the inside part of the human body through scanning the specific area. Pulmonary Segmentation of lungs from chest CT images is useful in diagnosis of abnormalities and surgery planning. In proposed method, Lung is segmented from chest CT using thresholding method and watershed transformation contributes to improve the numerical results for image segmentation. Further processing is carried out to visualize fissures characteristics and each segment inside the lung. Lung fissures are extracted using GLCM. The method is evaluated on chest CT images of normal and abnormal cases classification of Neural Network.

Keywords**—**Early Detection; HOG+GLCM Algorithm; Thresholding Method; Watershed Transformation.

Abbreviations**—**Computer Aided Diagnosing (CAD); Gray Level Co-occurrence Matrix (GLCM).

I. INTRODUCTION

LUNG cancer is a disease of abnormal cells multiplying and growing into a tumor. The mortality rate of lung cancer is the highest among all other types of cancer. Lung cancer is one of the most serious cancers in the world, with the smallest survival rate after the diagnosis, with a gradual increase in the number of deaths every year. Survival from lung cancer is directly related to its growth at its detection time. But people do have a higher chance of survival if the cancer can be detected in the early stages. Cancer cells can be carried away from the lungs in blood, or lymph fluid that surrounds lung tissue. Lung cancer can be divided into two main groups, non-small cell lung cancer and small cell lung cancer. As for the stages, in general there are four stages of lung cancer; I through IV. Staging is based on tumor size and tumor and lymph node location. Presently, CT are said to be more effective than plain chest x-ray in detecting and diagnosing the lung cancer [Seijion Shim & Jin Sam Kwak, 2].

II. EXISTING SYSTEM

There are three main processes used throughout the report: Pre-processing, feature extraction and finally the classification process. MATLAB is used in every process made throughout the project. Pre-processing step aims to reduce the noises in these images. Different filtering techniques were proposed in literature to remove these noises, such as median filtering, wiener filtering, Gaussian filter, bilateral filtering and a specific high-pass filter. Many others works combine median filters with Laplacian filters by a differential technique, which subtracts a nodule suppressed image (through a median filter) from a signal enhanced image (through a Laplacian matched filter with a spherical profile) difference image, containing nodule enhanced signal, is then obtained and used for the next stages [Chi Wan Sung & Kin Kwong Leung, 3]. An automated pulmonary nodule detection system mainly consists of two steps:

- Candidate screening and
- False positive reduction.

In candidate screening, a considerable number of coarse candidates are rapidly screened throughout the whole volume using a variety of criteria, e.g., intensity thresholding, shape curvedness and mathematical morphology. In false positive reduction, effective classifiers together with discriminative features are developed to reduce a large number of false positive candidates [Syed Hussain Ali & Victor C.M. Leung, 4].

In order to maintain a high sensitivity in candidate screening, the criteria employed in this step are usually quite straightforward and lenient, and consequently a great number of candidates are selected out and forwarded to the second step. In this regard, the false positive reduction stands as the most crucial component of an automated pulmonary nodule detection system and a lot of efforts have been dedicated to improving the performance of this step [Xiaorui Wang & Ming Chen, 5].

Disadvantages:

- Due to excessive over-segmentation the quality is low.
- Recognition of the image is not efficient

III. PROPOSED SYSTEM

3.1. System Architecture

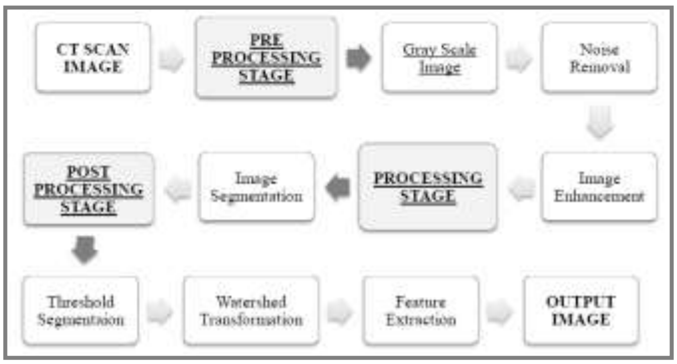

Figure 1: System Architecture

3.2. Segmentation

- Watershed algorithm.
- Threshold based method
- GLCM features

3.3. For Classification

Normal & abnormalities: Neural Network

The proposed methodology for detection and diagnosis of cancer detection from microscopic biopsy images consists of the stages of images enhancement, segmentation, feature extraction, and classification. The GLCM is a tabulation of how often different combinations of pixel brightness values (grey levels) occur in an image [Namyoon Lee et al., 6; Yaremko et al., 7]. The initial stage of the proposed Computer Aided Diagnosing (CAD) techniques is the extraction of lung region from the CT scan image. The lung CT images having low noise when compared to scan image and MRI image Firstly we create gray-level co-occurrence matrix from image using graycomatrix function in MATLAB. Gray level Co-occurrence matrix (GLCM) based texture feature extraction introduced by Haralick et al and Mari Partio et al., has been considered as the powerful technique and still now has been used in many applications of remote sensing for texture analysis [Wei Yu et al., 8].

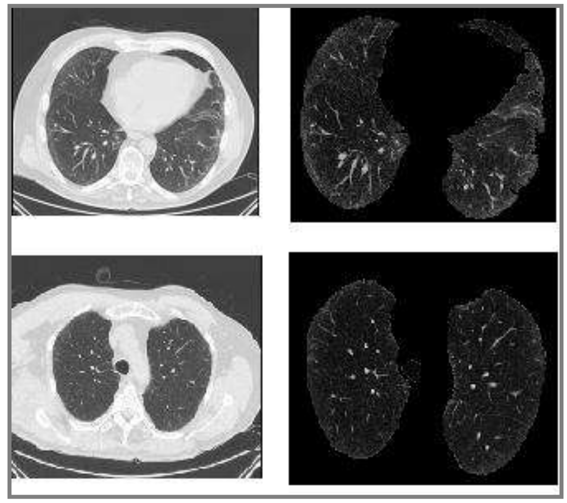

Figure 2: Some Results of the Proposed System

3.4. Proposed System for Lung Segmentation and Tumor Classification

A new system is proposed for lung segmentation and tumor classification from CT scan images.

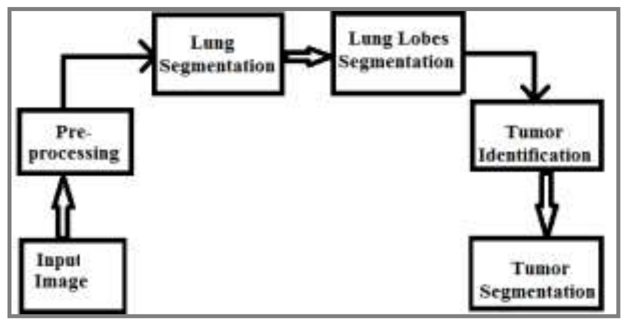

Figure 3: Proposed System for Lung Segmentation and Tumor Classification

3.4.1. Input Image

The input images are chest CT scan images in JPEG format that contain tumors. First image selected from the file specified by the string file name. The user has to select the required lung CT scan image for further processing. Then each image is re-sized to 256*256.The input image is in RGB format [ElGamal et al., 9]. So it is first converted into gray scale image for further processing.

Figure 4: Input CT Image of Lung

3.4.2. Wiener Filtering: (Pre-Processing)

The input image is in RGB format. So it is first converted into gray scale image for further processing [Lee et al., 10].

Then wiener filter of mask size 3*3 is used to remove noise because it is one of the best methods to remove the noise from the CT images, since these images usually contain artifacts or noise due to patient movement.

3.4.3. Lung Segmentation

In this module we segment left and right lung from the CT image. First we have chosen the seed point in the CT image. From the point we found intensity value of the image [Papadogiannis et al., 11]. We compare the intensity value between the neighboring pixels and current pixel. If the neighbor pixels values are related to the seed value, it will segment lungs from the original image. Threshold value between 0 and 180 is selected.

3.4.4. Lobe Segmentation

Watershed transformation is a common technique for image segmentation. However, its use for automatic medical image segmentation has been limited particularly due to over segmentation and sensitivity to noise. Employing prior shape knowledge has demonstrated robust improvements to medical image segmentation algorithms [Zhang & Dai, 12]. We propose a novel method for enhancing watershed segmentation by utilizing prior shape and appearance knowledge.

3.4.5. Tumor Identification

The feature vector is given as input to the classifier. This method differentiates and identifies the non-cancerous (Benign) and cancerous (Malignant) lung nodules [Lee et al., 13]. Classification and recognition randomly divide database into 70% of the database for training and 30% for testing. Both subset have the random samples from the same distribution.

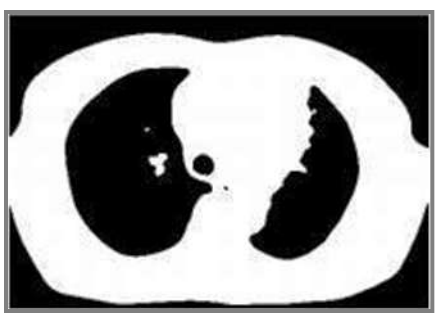

Figure 5: Output Image

3.4.6. Tumor Segmentation

Marker-controlled watershed segmentation follows this basic procedure and as Compute a segmentation function. This is an image whose dark regions are the objects you are trying to segment. 2) Compute foreground markers. These are connected blobs of pixels within each of the objects. 3) Compute background markers. These are pixels that are not part of any object. 4) Modify the segmentation function so that it only has minima at the foreground and background marker locations. 5) Compute the watershed transform of the modified segmentation function [Andrews et al., 14].

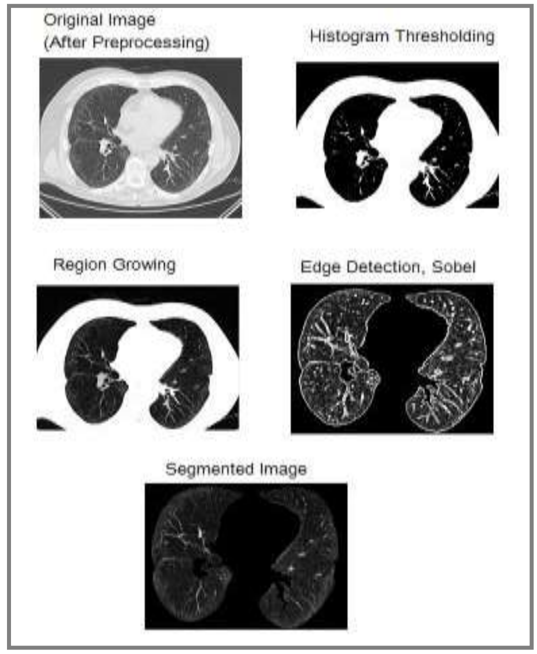

Figure 6: Result of Proposed System

3.4.7. Advantages

- High quality due to feature extraction
- Classifier is used to recognition the input in an efficient manner.

3.4.8. Applications

- Diagnosis the disease in the early stage itself

IV. Result and Discussion

The joint scheduling and power allocation is simulated for 40 nodes spread randomly in a 1500 * 1500m area network; transmission range for each node is random. Nodes are positioned randomly on the plane. Nodes start its travel from a random location to a random direction with a random speed. Nodes are arranged with three cluster heads as distribution of nodes selects one cluster head in a group before transmission of data occurs. The figure 7 shows the nodes indicated by different colours to show the difference between the cluster nodes that the data is transmitted from one node to another.

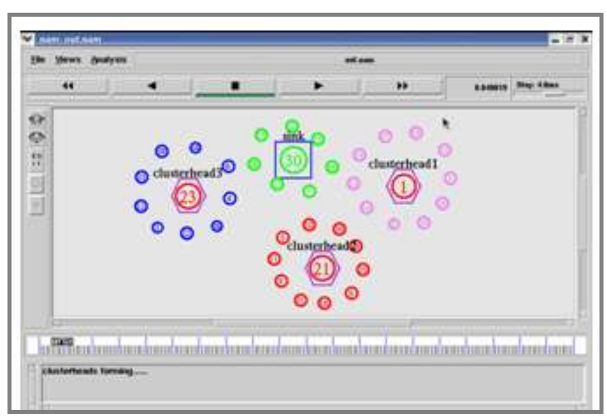

Figure 7: Nodes are arranged with Cluster Head after Data Transmission

The figure 7 shows the transmission of data for which the power is consumed. We can calculate the energy efficiency of data transmitted from source to destination. The nodes are denoted by different colours in the figure like pink, green, red i.e., just an identification. The figure 8 shows the cell edge throughput and the power consumed while message sending in the uplink and downlink coordinated system [Misun Yoon et al., 1]. The existing algorithm i.e., IW clustering algorithm improves 5% cell edge user rate. Therefore, the performance of cell edge users and power is controlled by using algorithm. The cell edge user rate is improved as 17% in our algorithm.

In downlink coordinated system, the power is highly consumed at the reception antenna and the total energy consumed to exchange a packet over a wireless during a single transmission and reception of packet. Power consumption depends on the packets send and received at the transmitter and reception respectively.

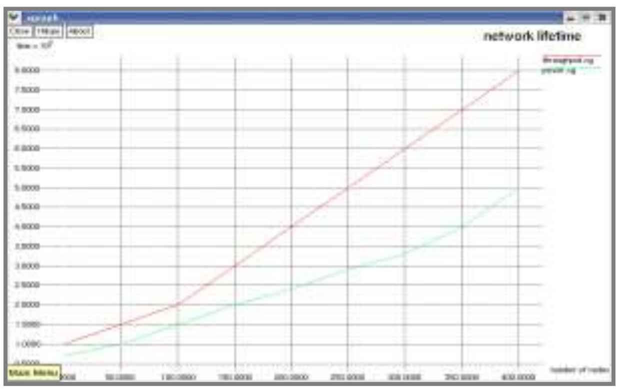

Figure 8: Graph for Throughput vs Power

The GSCA algorithm explains the static clustering algorithm do not guarantee the performance of the cell edge users and do not reflect the change of environment. The MAX-CG algorithm in figure 9 improves both the sum rate and the weak user' rate and it catches up with the performance of the FSCA. Never the less complexity of the MAX-CG Clustering algorithm is higher than the GSCA because it calculates the sum rate of all the combinations.

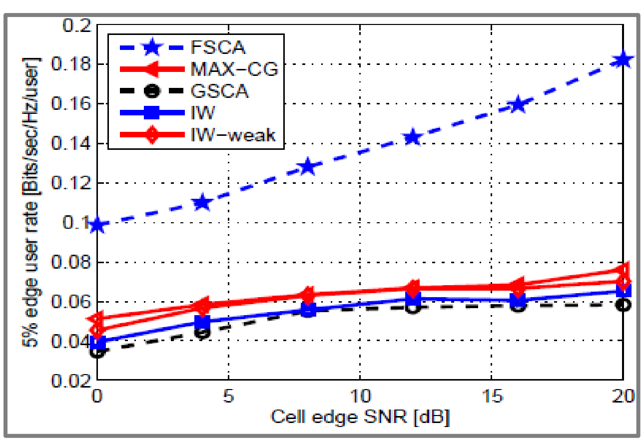

Figure 9: Cell Edge User Rate of FSCA, GSCA, MAX-CG, IW and IW-Weak

Thus, the cell edge user rate is improved as 17% without any handoff problem. At the same time, power is constraint to 5V so that the efficiency of the network lifetime is improved. The cell edge users in a mobile adhoc networks enables good signalling between transmitter and reception. The inter-cell interference is reduced by pre-scheduling the packet transmission from one node to another. The interference is assisted to the transmission for the cell edge users by joint scheduling and power control algorithm restricts the power for the packets movement by using proper channel and the output is displayed in the graph with comparison of existing algorithm.

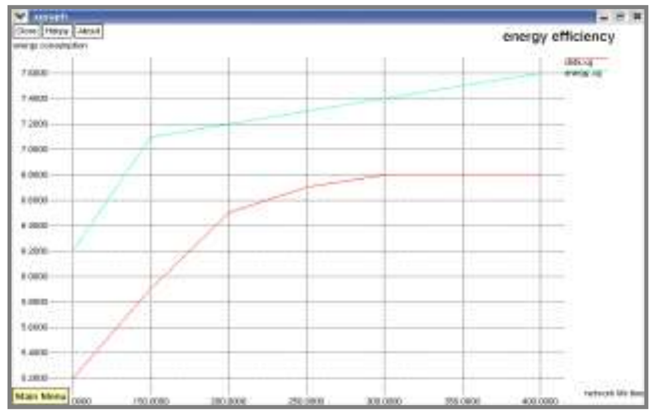

Figure 10: Graph for Data vs. Energy which improves the Network Lifetime and Power is Constraint to 5V when Data Transmitted

The real time application of this project, a radio cannot transmit to multiple radios at the same time, which means two links cannot share the same transmitter radio for simultaneous transmissions. The radio conflict is usually expressed as integer constraints in optimization problems or considered separately aside from other resource allocation. Since want to fully avoid integer components and facilitate linear programming towards a joint solution, generalize the definition of to cover not only co-channel interference but also the radio conflict relationship, which is defined as interference coefficient. Instead of preventing two links using the same radio for transmission, we allow the simultaneous transmissions but apply very large mutual interference between them by defining a large value for their interference coefficient.

Coordinated multipoint system is used in 4G advanced LTE system as relay technology where the technology suffer from reduced data rates at the cell edge users where signals levels are lower and interference levels are typically higher.

It is used in the Remote Radio units of the base station are located at the different location in space using Coordinated multipoint system act as a single unit behave as a intersite coordinated multipoint system. It used in the mobile computing, Radio communication, large business areas, military and medical networks.

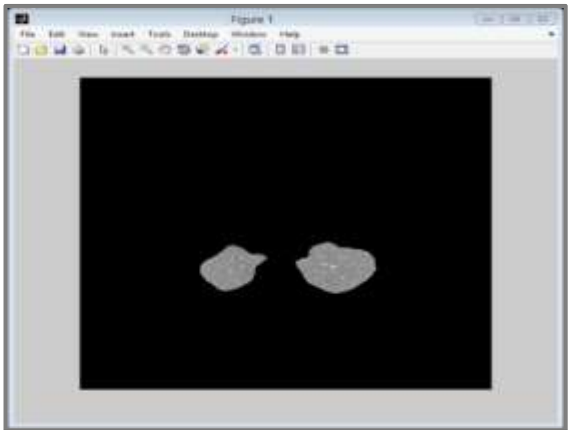

Figure 11: Normal Stage

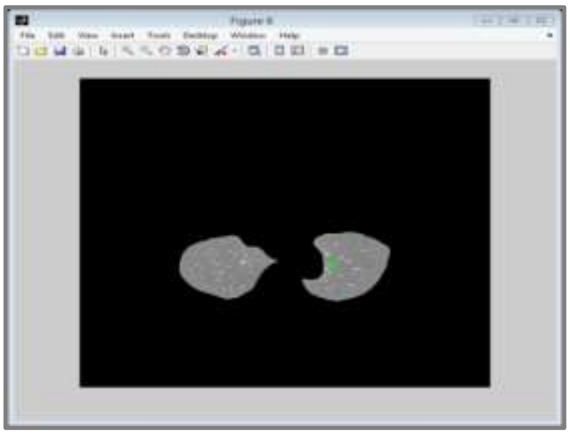

Figure 12: Benign Stage

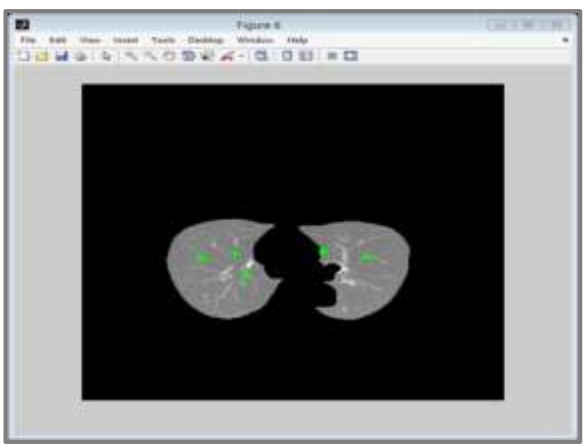

Figure 13: Malignant Stage

V. CONCLUSION

This paper is to find the cancerous cells present in the CT images of lung and give more accurate result by using various enhancement and segmentation techniques such as thresholding and watershed transform. The result shows that we can detect lung cancer at early stage thereby improving the survival time of affected patient. The automated identification of cancerous cells from microscopic images helps in alleviating the abovementioned issues and provide better result.

REFERENCES

[1] Misun Yoon, Myonk–seokkim & chungyong Lee (2015), "A Dynamic Cell Clustering Algorithm for Maximization of Coordination Gain in Uplink Coordinated System", *IEEE Transactions on Vehicular Technology*, Vol. 65, No. 3, Pp. 1752–1760.

[2] Seijion Shim & Jin Sam Kwak (2008), "Block Diagonalisation for Multiuser MIMO with Other Cell Interference", *IEEE Transactions on Wireless Communications*, Vol. 7, No. 7, Pp. 2671–2681.

[3] Chi Wan Sung & Kin Kwong Leung (2004), "Opportunistic Power Control for Throughput Maximization in Mobile Cellular System", *IEEE International Conference on Communications*, Vol. 5, Pp. 2954–2958.

[4] Syed Hussain Ali & Victor C.M. Leung (2009), "Dynamic Frequency Allocation in Fractional Frequency Reused OFDMA Networks", *IEEE Transactions on Wireless Communications*, Vol. 8, No. 8, Pp. 4286–4295.

[5] Xiaorui Wang & Ming Chen (2008), "Cluster-Level Feedback Power Control for Performance Optimization", *IEEE 14th International Symposium on High Performance Computer Architecture*, Pp. 101–110.

[6] Namyoon Lee, Xingqin Lin, Jeffrey G. Andrews & Robert W. Health Jr., (2015), "Power Control for D2D Underlaid Cellular Networks: Modelling, Algorithms and Analysis", *IEEE Journal on Selected Areas in Communications*, Vol. 33, No. 1, Pp. 1–13.

[7] O. Yaremko, B. Stryhalyuk, T. Maksymyuk, O. Lavriv & D. Kozhurav (2013), "Optimal Power Control Method in Multiuser Cellular Networks", *An International Quarterly Journal on Economics of Technology and Modelling Processes*, Vol. 2, No. 1, Pp. 63–67.

[8] Wei Yu, Taesoo Kwon & Changyong Shin (2013), "Multicell Coordination via Joint Scheduling Beamforming and Power Spectrum Adaptation", *IEEE Transactions on Wireless Communications*, Vol. 12, No. 7, Pp. 1–14.

[9] A. ElGamal, J. Mammen, B. Prabhakar & D. Shah (2006), "Optimal Throughput-Delay Scaling in Wireless Networks", *IEEE Transactions on Information Theory*, Vol. 52, No. 6, Pp. 2568–2592.

[10] T-H. Lee, J-C. Lin & T.S. Yu (1995), "Downlink Power Control Algorithms for Cellular Radio Systems", *IEEE Transactions on Vehicular Technology*, Vol. 44, No. 1, Pp. 89–94.

[11] A. Papadogiannis, D. Gesbert & E. Hardouin (2008), "A Dynamic Clustering Approach in Wireless Networks with Multi-Cell Cooperative Processing", *IEEE International Conference on Communications*, Pp. 4033–4037.

[12] H. Zhang & H. Dai (2004), "Cochannel Interference Mitigation and Cooperative in Downlink Multicell Multiuser MIMO Networks", *EURASIP Journal on Wireless Communications and Networking*, Vol. 2004, No. 2, Pp. 222–235.

[13] N. Lee, K. Lee, E. Nam, S. Kwon, J. Kang & G. Gil (2009), "Base Station Selection Technique for MMSE Joint Transmission in Downlink Cooperative MIMO System", *IEEE Wireless Communications and Networking Conference*, Pp. 1–5.

[14] J.G. Andrews, W. Choi & R.W. Health (2007), "Overcoming Interference in Spatial Multiplexing MIMO Cellular Networks", *IEEE Wireless Communications*, Vol. 14, No. 6, Pp. 95–104.