



Identification of Lung Cancer Related Genes Using Enhanced Floyd Warshall Algorithm in a Protein to Protein Interaction Network

Ashok Kasirayagouda Patil^{1*} Siddanagouda Somanagouda Patil² Prabhakar Manickam³

¹*Computer Science and Engineering, Sapthagiri College of Engineering, India*

²*Agril Statistics, Applied Mathematics & Computer Science, University of Agricultural Sciences, GKVK, India*

³*School of Computing and Information Technology, REVA University, India*

* Corresponding author's Email: akpatil.phd2018@gmail.com

Abstract: Lung cancer is defined as an uncontrolled cell growing in the tissues of lung, which is also said to be lung tumor. The lung cancer is curable in the starting stage, but identifying the lung cancer in starting stage is very difficult. In recent decades, researchers showed great interest on gene level lung cancer identification using shortest path between the lung cancer related genes. Many research has been done to identify the shortest path between the genes, but the conventional methods consumes more time for processing the data. In this research, Protein to Protein Interaction (PPI) structure is constructed from the weighted protein present in the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database. For identifying the shortest path between the genes in PPI, an effective algorithm: enhanced Floyd warshall algorithm is proposed. Floyd warshall is efficient in finding the shortest path between the genes and also solves all pairs of shortest path problem. A major drawback of Floyd warshall algorithm is, it works slower than other conventional algorithms designed to perform the same task. To improve the performance of traditional Floyd warshall algorithm, an iterative matrix is used for eliminating the invalid path. Then, the comparison between the proposed method and existing system is given in the experimental result. Experimental outcome shows that the proposed approach improved the time consumption up to 2-3 sec compared to the existing methods: Dijkstra's algorithm and Floyd warshall algorithm.

Keywords: Dijkstra's algorithm, Enhanced Floyd warshall algorithm, Protein to protein interaction, Search tool for the retrieval of interacting genes/proteins.

1. Introduction

In recent decades, lung cancer is one of the leading causes of cancer mortality [1]. Generally, primary lung cancer is characterized into two types such as, Small Cell Lung Cancer (SCLC) and Non-Small Cell Lung Cancer (NSCLC) [2]. Whereas, the most common types of NSCLC are squamous cell carcinoma, large cell carcinoma and adenocarcinoma. In the clinic, approximately 20% of lung cancer patients are diagnosed with SCLC and 80% are diagnosed with NSCLC [3, 4]. Hence, the biological behavior of SCLC and NSCLC are significantly different, these two distinct types of lung cancer grow in different ways and also treated differently. SCLC is highly malignant that is

characterized by rapid proliferation and metastasis [5]. Some NSCLC tumours grow and spread more slowly, making them less prone to developing early metastases and more amenable to surgical treatment during the early stages of the disease. Due to advancing experimental techniques, it was possible to research PPI [6, 7].

Currently, system biology is the most effective approach for understanding the molecular mechanisms of lung cancer [8]. Use of bio-informatics, enables the identification of unknown protein functions as well as new functions for familiar proteins based on the PPI analysis. Identifying the key nodes of proteins would be helpful in revealing the molecular mechanisms underlying lung cancer [9, 10]. In this experimental research, an enhanced Floyd warshall algorithm is

developed for predicting the patient susceptibility of lung cancer and to identify the diagnostic molecular markers for detecting early stages of lung cancer. Also, enhanced Floyd warshall algorithm helpful in finding the shortest path between each pair of genes. Whereas, all the genes present on the shortest paths were extracted and analysed, several of these genes have been related to lung cancer. These extracted lung cancer genes are helpful for bio-markers for diagnosis of lung cancer.

This literature is composed as follows. In Section 2, survey several recent PPI strategies. In section 3, enhanced Floyd warshall strategies is portrayed, to seek better shortest path between the genes. In Section 4, the execution of the implemented techniques are assessed by simulation. The conclusion is made in Section 5.

2. Literature review

Several researches are suggested by researchers in PPI based lung cancer detection. In this scenario, a brief evaluation of some important contributions to the existing literatures is presented.

B.Q. Li, J. You, L. Chen, J. Zhang, N. Zhang, H.P. Li, T. Huang, X.Y. Kong, and Y.D. Cai, [11] proposed a new method for identifying the lung cancer related gene in the PPI network. The PPI network was constructed based on the PPI data obtained from the STRING database. The 54 genes of the NSCLC and 84 genes of SCLC was retrieved based on the information from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. The shortest path between each pair of NSCLC and SCLC in the system was found using Dijkstra's algorithm. This literature showed that the proposed methodology was delivered a significant result in identifying the cancer gene. The proposed algorithm effectively identifies the shortest path between the genes, but identifying the related gene was one of the major drawback.

J. Zhang, B.Q. Li, T. Huang, L. Zhang, and Y.D. Cai, [12] presented a new method for identifying the retinoblastoma related genes using both expression profile and shortest path in the functional linkage graph. The construction of the PPI with weighted functionality linkage graph was done based on the data from STRING database. The 119 different genes between retinoblastoma and normal retina were obtained from the two gene expression studies of retinoblastoma. The result shows the efficiency of the proposed system and it was very useful in identifying the retinoblastoma related gene. This method identifies only a few retinoblastoma related

gene and also identifies other diseases related genes that were not specified in this literature.

L. Zhu, X. Chen, X. Kong, and Y.D. Cai, [13] proposed a computational method to identify the genes related to the Hepatitis C Virus (HCV) and trace element metabolism factor. The searching process were involved in three steps, (1) applying shortest path in the PPI network, (2) excluded the un-related genes in the path, and (3) find the core genes related to the HCV and trace element metabolism factor. Here 12 genes were obtained related to the 12 trace element metabolism process and it was associated with the HCV process. The genes were quite proved to be related to the HCV and trace element by the analysis, but still it consumes more time for identifying the cancer genes.

R. Sable, and S. Jois, [14] evaluated a docking methodology in PPI design. For therapeutic purposes, only a few methods were succeeding in the combination of both experimental and docking methods with scoring function. The docking algorithm was used in the early stage of PPI and a lot of data was available for docking, so it produces an efficient result in PPI. The shortest path between the cancer genes were identified, in that a few genes were irrelevant to diseases or cancer gene.

N. Dounghan, W. Engchuan, J.H. Chan, and A. Meechai, [15] illustrated the two search algorithms for Gene Sub-Network-Based Feature Selection (GSNFS) for subnetwork expansion with the degree of connectivity and scoring scheme for building an effective subnet works and topology. For each iteration, the neighbourhood genes of current subnet work were recruited. The proposed search algorithm provides the higher number of genes in the subnetwork expansion compared to the greedy algorithm. So, an effective search algorithm was required to measure the shortest path between the genes.

To overcome the above mentioned draw backs and for improving the shortest path between the cancer genes and to reduce the elapsed time, an effective methodology named as enhanced Floyd warshall strategy is implemented.

3. Proposed methodology

The shortest path of the genes is calculated by using enhanced Floyd Marshall Algorithm in PPI network. The PPI is constructed with the cancer related genes and their weight value is obtained from the STRING database. The STRING database consists of the large number of genes data and their respective disease for the various living organism.

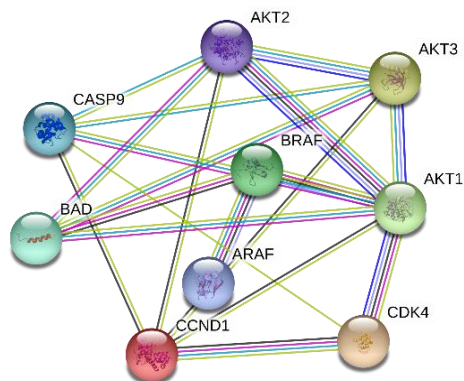


Figure.1 PPI structure of the given gene

The genes responsible for the SCLC and NSCLC are taken from the KEGG database. The shortest path for the genes are calculated effectively with the enhanced Floyd Marshall algorithm. This value can be used to identify the cancer in the gene level and it can found before the critical state. In the early stage, the lung cancer is curable by surgery and chemotherapy. The step by step procedure of proposed methodology is detailed below.

Step 1: Acquisition of lung cancer dataset from STRING database.

Step 2: In STRING database, KEGG pathway is used to refer the PPI structure.

Step 3: Then, classify the SCLC and NSCLC genes from the STRING database.

Step 4: Finally, identifying the shortest path between the genes (related to SCLC and NSCLC) in PPI using enhanced Floyd Warshall algorithm.

3.1 Dataset collection

We obtained STRING biological database from (SCRIBD 10.1.1.154 gene bank name) with the predicted PPI connection. The latest version of the STRING 10.0 contains the information for the 9.6 million protein about 2000 organism. The data are collected from the several source contains the experimental information, computational prediction and public text collection. This data can be freely accessible and this data can be used to highlight the functional enrichment based on the user listed proteins using the number of classification system such as GO, Pfam and KEGG. The graphical representation of PPI structure for gene is represented in the Fig. 1.

The STRING have the large amount of information about the many proteins for the various organisms. The user can obtain the information about the gene by including the name in the database. It can also highlight the functional enrichment of genes from the user provided the gene

from the number of classification system such as GO, Pfam and KEGG.

The genes are provided in the STRING database and information about the genes are obtained. The weighted values of the genes are used to plot the weighted graph and gene structure is shown in Fig. 1. The genes that are the major relevant to the small cell lung cancer and non-small cell lung cancer are taken. PPI network is the important factor to identify the shortest path in the gene. These networks are used for the filtering and assenting the active genomics gene, this provide the platform for annotating structural, functional and evolutionary properties of proteins. Exploring the prediction of the PPI interaction gives the new direction of the future experimental research and it also gives the cross species, which can use for the effective interaction mapping. STRING is the database that provide the information about the protein and it imports the data only for the experimentally derived PPI through literature curation. The STRING also stores the predicted PPI from (i) text mining of scientific texts, (ii) interactions computed from genomic features, and (iii) interaction transferred from model organisms based on orthology. The KEGG suggest a common reference of partnership that can be used to refer these PPI structure.

The data from STRING are weighted, integrated and a confidence score is obtained for all the protein interactions. There are two modes of data are present in the STRING, they are (1) protein-mode and the COG-mode. The interaction between the proteins are propagated for which interaction has been describe by the orthology. The result of the various computational prediction is analysed from the different views. The interface facility is also available in the STRING to access the data and to give the overview of the protein and their interactions. A plug-in for the cytoscape is also available in STRING and another possible way to access the STRING data is use the application by URL address that contain the request. The data has been collected from the STRING and PPI network is constructed. From which shortest path has been calculated by the Floyd warshall algorithm.

3.2 KEGG database

KEGG is a collection of database contains information regarding with genomes, biological pathways etc. KEGG is used for the bioinformatics researches and education, include data analysis and also for modelling and simulation in system biology. KEGG pathway clearly determines the gene responsible for the NSCLC are AKT1, BRAF,

AKT2, BAD, AKT3, CASP9, CCND1 and CDK4 and the gene responsible for the SCLC are AKT1, CASP9, AKT2, CCND1, AKT3 and CDK4.

3.3 Floyd warshall algorithm

Floyd algorithm is used to find the shortest path in a weighted graph, which is the minimum sum of weighted edges for all paths between the pairs. All pair's shortest path concern is to find the minimum weight path between any two vertices in the graph. The distant matrix delivers the edge weight for adjacent vertices. If the two vertices are adjacent then the corresponding distant matrix entry is ∞ . In addition, warshall procedure utilizes the adjacency matrix to identify the transitive closure of a directed graph.

The combined Floyd and warshall algorithm is used to find the shortest path in the system either it contains positive or negative value. The single cycle of the algorithm will find the shortest path in the vertices. The Floyd warshall algorithm is applied to the directed graph of the PPI structure, which is constructed from the data obtain from the STRING database. Initially, consider the given PPI structure as weighted graph $G = (V, E)$ with vertices V numbered from 1 to N and the edge E . The weight of an edge between the vertex i and vertex j in the graph $G = (V, E)$ is calculated by using Eq. (1).

$$w_{ij} = \begin{cases} 0 & \text{if } i = j \\ \text{weight of a directed edge}(i, j) & \text{if } i \neq j \text{ and } (i, j) \in E \\ \infty & \text{if } i \neq j \text{ and } (i, j) \notin E \end{cases} \quad (1)$$

The $n \times n$ matrix representing the edge weights of an n -vertex graph is denoted as $W = w_{ij}$. The distance of the shortest path from the vertices i and j with the intermediate vertex k in the set of the system $\{1, 2, 3, \dots, k\}$ is denoted as $d_{ij}^{(k)}$. In each pair of the vertices, there are two possible ways in the vertices (i) K is not a vertex on the path, so the shortest path has length $d_{ij}^{(k-1)}$ (ii) K is a vertex on the path, so the shortest path has length $d_{ik}^{(k-1)} + d_{kj}^{(k-1)}$.

Hence, the shortest path $d_{ij}^{(k)}$ is defined using recursion formula, which is denoted in Eq. (2).

$$d_{ij}^{(k)} = \begin{cases} w_{ij} & \text{if } k = 0 \\ \min(d_{ij}^{(k-1)}, d_{ik}^{(k-1)} + d_{kj}^{(k-1)}) & \text{if } k \geq 1 \end{cases} \quad (2)$$

Then, the predecessor of vertex j on a shortest path from vertex i with all intermediate vertices $\{1, 2, 3, \dots, k\}$ is given in Eqs. (3) and (4).

$$\pi_{ij}^{(0)} = \begin{cases} nil & \text{if } i = j \text{ or } w_{ij} = \infty \\ i & \text{if } i \neq j \text{ or } w_{ij} < \infty \end{cases} \quad (3)$$

$$\pi_{ij}^{(k)} = \begin{cases} \pi_{ij}^{(k-1)} & \text{if } d_{ij}^{(k-1)} \leq d_{ik}^{(k-1)} + d_{kj}^{(k-1)} \\ \pi_{kj}^{(k-1)} & \text{if } d_{ij}^{(k-1)} > d_{ik}^{(k-1)} + d_{kj}^{(k-1)} \end{cases} \quad (4)$$

The Floyd warshall algorithm first applies the shortest path (i, j, k) function to the vertices i to j when first set $k = 1$, then $k = 2$, etc., and this goes until $k = N$. The shortest path for all the (i, j) pair using the intermediate value.

3.4 Enhanced Floyd warshall algorithm

Enhanced Floyd warshall algorithm analysis the all possible path between the given pair of vertices. The Enhanced Floyd warshall algorithm possible to reconstruct the path with the simple modification, an iterative matrix is used in this algorithm for eliminating the invalid path. The path between the any two vertices have another similar path with same weight W in the reverse direction between the same two vertices. If there is no direct acknowledgement path or reverse path between the two vertices, then there is no direct path existing between the vertices. In case, if weight of the path is not equal to the reverse path then the respective path is invalid.

Consider the PPI structure as graph $G = (V, E)$. Before calculating the shortest path distance $d_{ij}^{(k)}$ from the vertices i and j with the intermediate vertices. Validate the path in condition presence of its corresponding acknowledgement path and also validated the path in which length is equal to the reply path. Using this two conditions, eliminate the invalid paths between the vertices $d_{ij}^{(l)}$. Then, the shortest path between the vertices i and j is identified for valid paths $d_{ij}^{(V)}$. The mathematical representation for eliminating the invalid paths is stated in Eq. (5).

$$d_{ij}^{(k)} = \begin{cases} d_{ij}^{(l)} & d_{ij} \neq d_{ji}, Wd_{ij} \neq Wd_{ji} \\ d_{ij}^{(V)} & d_{ij} = d_{ji}, Wd_{ij} = Wd_{ji} \end{cases} \quad (5)$$

After eliminating the invalid paths, the shortest path is identified among the valid paths $d_{ij}^{(V)}$ using Eq. (6).

$$d_{ij}^{(k)} = \begin{cases} \min(d_{ij}^{(k-1)}, d_{ik}^{(k-1)} + d_{kj}^{(k-1)}) & \text{if } k \geq 1 \\ w_{ij} & \text{if } k = 0 \end{cases} \quad (6)$$

The pseudo code of proposed methodology: enhanced Floyd warshall algorithm is detailed below.

3.4.1. Pseudo code of proposed algorithm

In this section, the pseudo code for enhanced Floyd warshall algorithm is presented. Compared with the conventional Floyd warshall algorithm, the enhanced Floyd warshall eliminates the invalid or unwanted paths. A few changes are carried out in the step 8. If the path is invalid, the condition automatically gets break. Finally, the loop ends after calculating the shortest path between the vertices using the steps 11 and 12.

1. Let distance be a $|V| \times |V|$ array of minimum distances initialized to ∞ (infinity)
2. For each weighted graph $G = (V, E)$, $E \leftarrow$ edge
3. $dist[i, j] \leftarrow w(i, j)$ // weights of an N-vertex graph
4. For each vertex V
5. $dist[V][V] \leftarrow 0$
6. For k from 1 to $|V|$
7. For i from 1 to $|V|$
8. If ($dist[i][k] \neq \infty$)
9. For j from 1 to $|V|$
10. If ($dist[k][j] \neq \infty$)
11. If ($dist[i][j] > dist[i][k] + dist[k][j]$)
12. $dist[i][j] \leftarrow dist[i][k] + dist[k][j]$
13. End
14. End
15. End
16. End
17. End
18. End

4. Experimental result

For experimental simulation, MATLAB (version 2017a) was employed on PC with 3.2 GHz with i5 processor. MATLAB run on operating system of windows 7/8/10, Linux, Mac with processor of Intel and AMD x86 processor. In this experimental research, the Genes that were undertaken for analysing NSCLC and SCLC were AKT1, AKT2, AKT3, ARAF, BAD, BRAF, CASP9, CCND1 and CDK4.

4.1 Non-small cell lung cancer

Lung cancer is a leading cause of cancer death among men and women in industrialized countries.

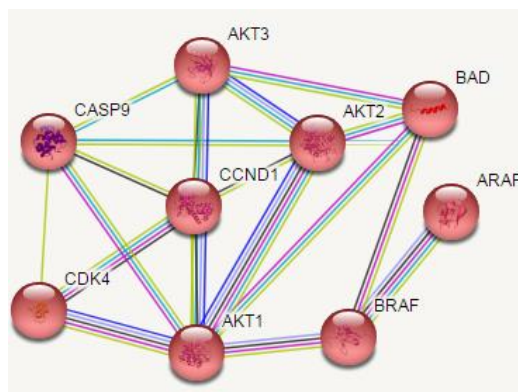


Figure.2 Gene representation for NSCLC

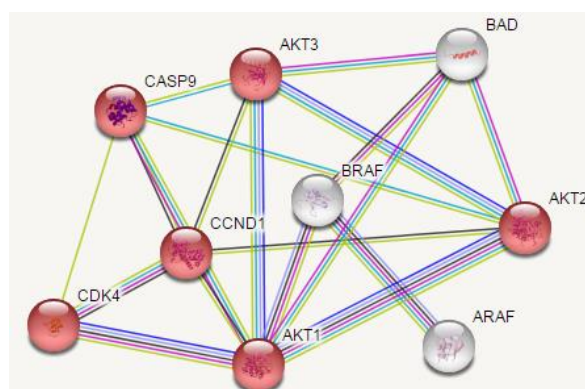


Figure.3 Gene representation for SCLC

Lung cancer is generally divided into two types such as, NSCLC and SCLC, these types of lung cancer are very different to treat. NSCLC accounts approximately 85% of lung cancer and represents the gene that are responsible for cancer, mainly consists of squamous cell (SCC), adeno (AC) and large-cell carcinoma. The graphical representation of NSCLC is represented in Fig. 2.

4.2 Small cell lung cancer

Lung cancer occurs mainly in current or former smokers, also it is the most common type of diseases seen in non-smokers. In Fig. 3, SCLC is a highly aggressive neoplasm, which accounts approximately 25% of all lung cancer cases. The graphical representation of SCLC is represented in the Fig. 3.

Where, red gene is represented as a query protein and first shell of interactors, and the white gene is stated as a second shell of interactors. The comparison of Dijkstra's algorithm, Floyd-Warshall algorithm and Enhanced Floyd-Warshall algorithms are described in the Table 1 by means of elapsed time (sec). Compared to Dijkstra's algorithm and Floyd-Warshall algorithm, the enhanced Floyd-Warshall algorithm is faster, because the process for finding the destination vertices through the other vertices has been removed.

Table 1. Performance comparison of existing and proposed methodologies of time complexity

Number of Nodes	Time (In Second)		
	Dijkstra's	Floyd-Warshall	Enhanced Floyd-Warshall
19	1.5729	0.60001	0.44184
31	1.88	0.83372	0.6622
41	2.2405	1.0149	0.87155
49	4.5843	1.2306	1.0745
58	5.0149	1.4522	1.2788
60	5.1125	1.5314	1.378
65	5.4822	1.7113	1.5618

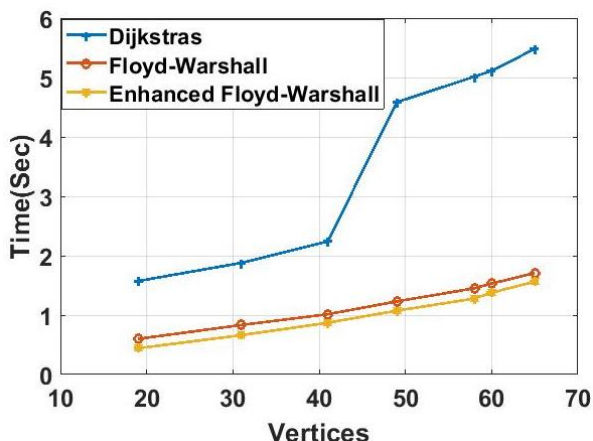


Figure. 4 Graphical comparisons of existing and proposed methodologies of time complexity

This process is removed due to the shortest path between the direction is only used for the analysis.

All gene nodes are taken from the STRING dataset, which fed into the Matlab and plotted. The shortest path between the genes of SCLC and NSCLC are plotted by using Enhanced Floyd-Warshall algorithm. The shortest path is obtained by calculating the path between the AKT1 and CCND1 gene. The interaction between the genes are obtained by selecting the nodes and genes belong to the SCLC and NSCLC. There are 2 stages in building a routing table in Link-State Routing (LSR) protocols. First, the map of the entire network is stored in every router, and then the shortest distance of each node is calculated by using each router.

Fig. 4 represents the comparison of Dijkstra's algorithm, Floyd-Warshall algorithm and enhanced Floyd-Warshall algorithm in "vertices vs time". Compared to the existing methods (Dijkstra's algorithm and Floyd-Warshall algorithm), the Enhanced Floyd-Warshall algorithm is twice significant in time consumption. Enhanced Floyd-warshall algorithm is used to find the shortest path among all pairs of nodes, which does not contain any cycles of negative length. The main advantage of enhanced Floyd-Warshall algorithm is its simplicity. The graphical representation of shortest

path between the AKT1 and CCND1 gene for SCLC is shown in Fig. 5.

In Fig. 6, the shortest path between the genes of NSCLC is plotted using enhanced Floyd-Warshall algorithm. It is also obtained by determining the shortest path between the AKT1 and CDK4 as shown in Fig. 6. In NSCLC, the interaction between the genes are achieved by selecting the nodes and genes, which belongs to the NSCLC.

4.3 Comparative analysis

Table 2 presents the comparative study of existing work and proposed work performance. B.Q. Li, J. You, L. Chen, J. Zhang, N. Zhang, H.P. Li, T. Huang, X.Y. Kong, and Y.D. Cai, [11] proposed a computational method for identifying lung-cancer related genes with a shortest path approach (Dijkstra's algorithm) in a PPI network. Based on the PPI data from STRING, a weighted PPI network was constructed with 54 NSCLC and 84 SCLC-related genes were retrieved from associated KEGG pathways. This experiment was carried out on a publicly available database (i.e., STRING database) to validate its result in terms of time and distance. For instance, the existing work consumed 5.4822 sec for finding the shortest path between the 65 nodes.

Similarly, F. Towfic, S. VanderPias, C.A. Oliver, O. Couture, C.K. Tuggle, M.H.W. Greenlee, and V. Honavar, [16] presented a detection method: Floyd warshall algorithm for finding genes that participate in similar biological processes across different organisms, also extending the interactions between genes across different pathways, and understanding the evolution of gene families. This work consumed 1.7113 sec for finding the shortest path between the 65 nodes. Whereas, the proposed work took only 1.5618 Sec of time that was lower than the existing work (Dijkstra's algorithm). Table 2 confirmed that the proposed work performed effectively than the existing shortest path distance measure methods.

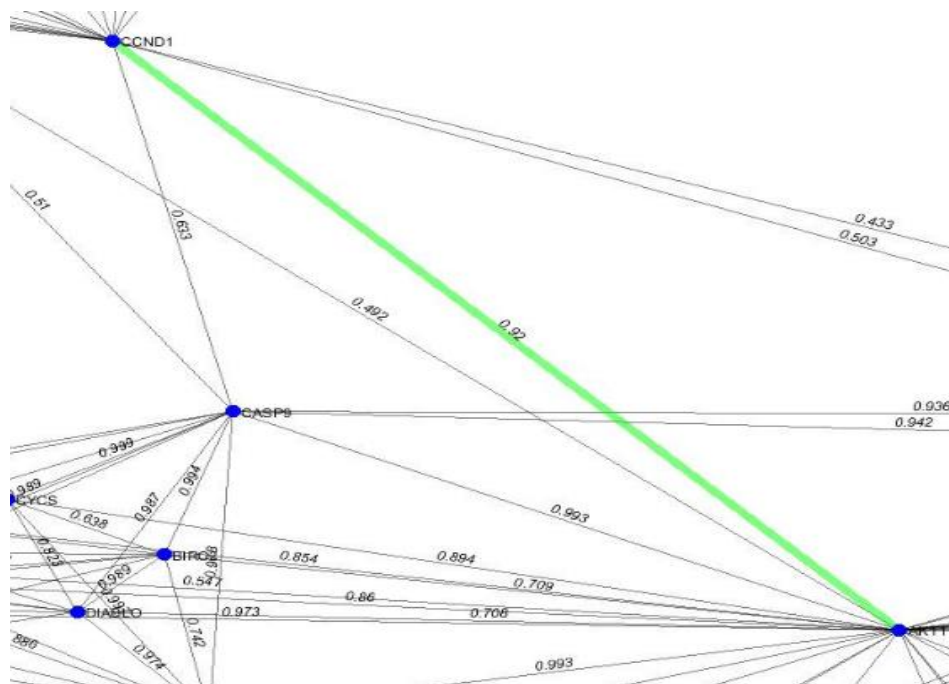


Figure. 5 Shortest path representation of SCLC gene

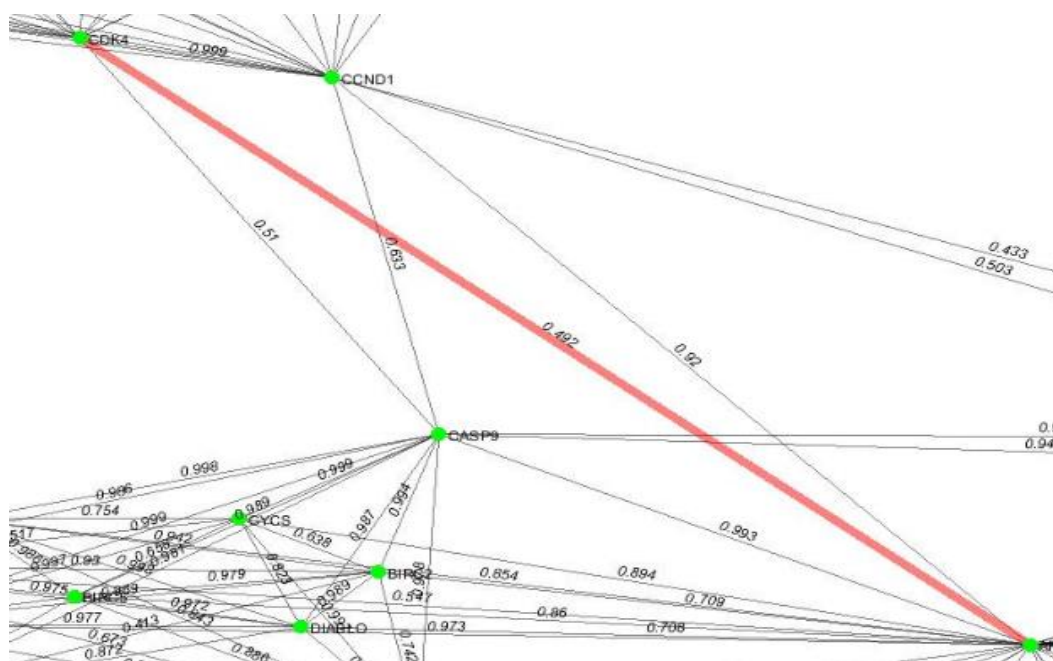


Figure. 6 Shortest path representation of NSCLC gene

Table 2. Performance comparison

Number of nodes=65			
Methods	Time (sec)	Distance	
		SCLC	NSCLC
Dijkstra's algorithm [11]	5.4822	0.92	0.492
Floyd-Warshall algorithm [16]	1.7113	0.92	0.492
Enhanced Floyd-Warshall algorithm	1.5618	0.92	0.492

5. Conclusion

In the recent decades, lung cancer is one of the common type of cancer present among the humans. The initial stage of the lung cancer is found by identifying the shortest path between the genes related to the lung cancer. The NSCLC and SCLC are the major cancer gene present in the lung cancer. In this paper, an enhanced Floyd warshall algorithm is used to find the shortest between the genes by

eliminating the invalid paths, which are responsible for the lung cancer. PPI network is constructed from the data obtained from STRING database and KEGG pathway. The KEGG pathway gives the genes responsible for the NSCLC and SCLC. Associated with the other obtainable approaches, the proposed scheme delivered an effective performance by means of elapsed time, around 2-3 (Sec) time consumption than the previous methods. In the future work, for further improving the elapsed time rate, a combination of distance based strategy is developed.

References

- [1] R. Enyinnaya, "Predicting Cancer-Related Proteins in Protein-Protein Interaction Networks using Network Approach and SMO-SVM Algorithm", *International Journal of Computer Applications*, Vol.115, No.3, pp.5-9, 2015.
- [2] A. Grossmann, N. Benlasfer, P. Birth, A. Hegele, F. Wachsmuth, L. Apelt, and U. Stelzl, "Phospho-tyrosine dependent protein-protein interaction network", *Molecular systems biology*, Vol.11, No.3, pp.794, 2015.
- [3] H. Xu, J. Ma, J. Wu, L. Chen, F. Sun, C. Qu, D. Zheng, and S. Xu, "Gene expression profiling analysis of lung adenocarcinoma", *Brazilian Journal of Medical and Biological Research*, Vol.49, No.3, 2016.
- [4] C.H. Wu, C.L. Hsu, P.C. Lu, W.C. Lin, H.F. Juan, and H.C. Huang, "Identification of lncRNA functions in lung cancer based on associated protein-protein interaction modules", *Scientific reports*, Vol.6, No.5, 2016.
- [5] W. Yu, L.R. He, Y.C. Zhao, M.H. Chan, M. Zhang, and M. He, "Dynamic protein-protein interaction subnetworks of lung cancer in cases with smoking history", *Chinese Journal of Cancer*, Vol.32, No.2, pp.84, 2013.
- [6] S. Zhang, C.C. Liu, W. Li, H. Shen, P.W. Laird, and X.J. Zhou, "Discovery of multi-dimensional modules by integrative analysis of cancer genomic data", *Nucleic acids research*, Vol.40, No.19, pp.9379-9391, 2012.
- [7] F. Fontaine, J. Overman, and M. François, "Pharmacological manipulation of transcription factor protein-protein interactions: opportunities and obstacles", *Cell Regeneration*, Vol.4, No.1, pp.2, 2015.
- [8] M. Arkin, "Protein-protein interactions and cancer: small molecules going in for the kill", *Current Opinion in Chemical Biology*, Vol.9, No.3, pp.317-324, 2005.
- [9] J. Zhang, T.D. Le, L. Liu, J. He, and J. Li, "A novel framework for inferring condition-specific TF and miRNA co-regulation of protein-protein interactions", *Gene*, Vol.577, No.1, pp.55-64, 2016.
- [10] K. Anbarasu and J. S. Kumar, "Multidimensional significance of crystallin protein-protein interactions and their implications in various human diseases", *Biochimica et Biophysica Acta (BBA)-General Subjects*, Vol.1860, No.1, pp.222-233, 2016.
- [11] B.Q. Li, J. You, L. Chen, J. Zhang, N. Zhang, H.P. Li, T. Huang, X.Y. Kong, and Y.D. Cai, "Identification of lung-cancer-related genes with the shortest path approach in a protein-protein interaction network", *BioMed research international*, 2013.
- [12] J. Zhang, B.Q. Li, T. Huang, L. Zhang, and Y.D. Cai, "Identification of retinoblastoma related genes with shortest path in a protein-protein interaction network", *Biochimie*, Vol.94, No.9, pp.1910-1917, 2012.
- [13] L. Zhu, X. Chen, X. Kong, and Y.D. Cai, "Investigation of the roles of trace elements during hepatitis C virus infection using protein-protein interactions and a shortest path algorithm", *Biochimica et Biophysica Acta (BBA)-General Subjects*, Vol.1860, No.11, pp.2756-2768, 2016.
- [14] R. Sable and S. Jois, "Surfing the protein-protein interaction surface using docking methods: application to the design of PPI inhibitors", *Molecules*, Vol.20, No.6, pp.11569-11603, 2015.
- [15] N. Doungpan, W. Engchuan, J.H. Chan, and A. Meechai, "GSNFS: Gene subnetwork biomarker identification of lung cancer expression data", *BMC medical genomics*, Vol.9, No.3, pp.70, 2016.
- [16] F. Towfic, S. VanderPlas, C.A. Oliver, O. Couture, C.K. Tuggle, M.H.W. Greenlee, and V. Honavar, "Detection of gene orthology from gene co-expression and protein interaction networks", *In BMC bioinformatics, BioMed Central*, Vol. 11, No. 3, pp. S7, 2010.