

Detection of Prostate Cancer Related Genes using Modified Ford-Fulkerson Algorithm in Protein-to-Protein Interaction Network



Sanjeev Prakashrao Kaulgud, Vishwanath Hulipalled, Siddanagouda Somanagouda Patil

Abstract: Prostate cancer is a malignancy cancer that affects prostate epithelial cells. Presently, prostate cancer is the second leading cause of cancer-related death in men. In this research, a new computational system was proposed for determining the prostate cancer related genes with the shortest path methodology in a Protein to Protein Interaction (PPI) network. Here, a weighted PPI network was constructed on the basis of PPI data from Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database. Totally, eighteen prostate related genes were extracted from the STRING database by using Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway. Then, the shortest path between eighteen genes was identified using modified Ford-Fulkerson algorithm. Generally, the conventional Ford-Fulkerson algorithm was very effective in detecting the shortest path between the prostate cancer-related genes, but the elapsed time was high when the PPI network has more number of genes. In order to reduce the elapsed time, the modified Ford-Fulkerson algorithm was developed by eliminating the invalid path in gene connection. In the experimental section, the proposed shortest path approach reduced the elapsed time up to 0.025-0.002 seconds compared to the existing shortest path methodologies.

Index Terms: Modified Ford-Fulkerson algorithm, prostate cancer, protein to protein interaction, and shortest path.

I. INTRODUCTION

Prostate cancer is the second most common cancer among men and it is currently diagnosed around 10,000 men each year and accounts for approximately 2,500 deaths per year. Men with advanced prostate cancer have symptoms like poor stream, lymphoedema, bone pain, etc., and also have systematic advanced cancer prostate symptoms like lethargy, cachexia, and weight loss [1-2]. The main staging and diagnostic investigations utilized for prostate cancer are magnetic resonance imaging, prostate-specific antigen elevation, computed tomography, digital rectal examination,

transrectal ultrasonic sound, X-rays, etc. The success of treatment in early-stage prostate cancer involves in balancing the disease aspects against the patient quality of life and also depends on the physician's ability in responding to the desires and interests of the individual [3-4]. Generally, the elder patients have some form of hormonal therapy, which is treated with anti-androgen monotherapy [5-6]. As prostate cancer regularly needs the testosterone (male sex hormone) for its growth that is the first line therapy for lowering the serum-testosterone, which is done by either medical or surgical castration. The main types of prostate cancers are small cell prostate and non-small cell prostate [7-8].

Nowadays, system biology (PPI) plays an effective role in understanding the molecular mechanisms of prostate cancer. The bio-informatics helps to identify the functions of familiar and unknown proteins [9-10]. Determining the familiar and unknown proteins reveals the molecular mechanisms underlying prostate cancer. The current research aimed to determine the potential genes related to prostate cancer for enhancing the prediction of prognosis of patients with prostate cancer. In this research paper, a modified Ford-Fulkerson algorithm was proposed to determine the diagnostic molecular markers for predicting the early stages of prostate cancer and also to identify the patient susceptibility of prostate cancer. Additionally, the modified Ford-Fulkerson algorithm was very useful for determining the shortest between each pair of prostate related genes (18 prostate genes; BCL2, CCND1, CCNE1, CCNE2, CDKN1A, CDKN1B, CHUK, CREBBP, E2F1, IKBKB, IKBKG, MDM2, NFKB1, NFKBIA, PTEN, RB1, RELA, and TP53). In this scenario, a weighted PPI network was constructed with PPI data, which were extracted from the STRING database using KEGG pathway. The analysis of shortest path between prostate related cancer genes helps the bio-markers for earlier diagnosis of prostate cancer.

This research paper is prepared as follows. Several recent papers in the PPI network for prostate cancer detection are detailed in section II. Detailed explanation about proposed methodology is given in section III. Section IV illustrates the quantitative analysis of proposed methodology. Conclusion is made in section V.

II. RELATED WORK

Numerous methodologies are developed by the researchers in PPI network for prostate cancer detection.

Revised Manuscript Received on October 30, 2019.

* Correspondence Author

Sanjeev Prakashrao Kaulgud, Department of Computer Science & Engineering, REVA and Presidency University, Bengaluru, India.

Vishwanath Hulipalled, School of Computing and Information Technology, REVA University, Bengaluru, India.

Siddanagouda Somanagouda Patil, Department of Agri Stat, Applied Maths & Computer Science, University of Agricultural Sciences, Bengaluru, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Detection of Prostate Cancer Related Genes using Modified Ford-Fulkerson Algorithm in Protein-to-Protein Interaction Network

In this sub-section, evaluation of a few essential contributions to the existing literatures are presented. S. Detchokul, E.J. Crampin, A. Elangovan, A.G. Frauman, and M.J. Davis, [11] developed an androgen independent sub-line from LNCaP cells by prolonged disclosure to androgen-deprivation. The developed methodology investigated the phenotypic profiles by performing RNA-sequencing. The generated reads were associated to human clinical samples, which were evaluated by performing PPI, pathway analysis and differential expression. The pathway analysis exposed the same patterns of enriched pathways from differentially expressed genes of both cell line and human clinical datasets. This pathway analysis revealed many network interactions and potential mechanisms that includes co-operative behavior of nuclear receptors. The relationship between the progesterone receptors does not compel, when androgen receptor was not expressed in network analysis.

L. Chen, Z. Zhang, Y.H. Zhang, Y.D. Cai, and T. Huang [12] combined two network diffusion approaches; Random Walk with Restart (RWR), and Laplacian Heat Diffusion (LHD) for identifying the essential genes in the entire network. The developed system carried-out three screening tests for increasing the reliability of putative genes. After evaluating the putative genes, the developed system assumed twenty-nine genes, which were selected by LHD approach. Then, RWR approach was used for identifying the normal and abnormal genes. Additionally, a few genes have special positions that were selected by RWR or LHD algorithms. Still, the evaluated genes have no associations with biological processes for protecting cells form the malignant alternations.

Y. Lin, L. Shen, F. Chen, X. Tang, Z. Sun, C. Du, J. Chen, H. Ding, and B. Shen, [13] predicted the prostate cancer metastasis on the basis of network vulnerability analysis by identifying key microRNAs as biomarkers. The developed system extracted the mRNAs and microRNAs, which were expressed between Metastatic Prostate Cancer (MPCa) samples and primary prostate cancer. Then, a new bioinformatics approach was developed to construct screened microRNA biomarkers and MPCa-specific microRNA-mRNA network. The developed approach highlighted the network vulnerability and system stability changes with three measurements; importance of microRNA targets, regulation of single line structure and the percentage of transcription genes in microRNA targets. In this research study, the developed approach treated the genes equally, but the essential genes were varied in different biological activities, so more functional annotations need to be weighted.

L.C. Lin, C.H. Lai, A.C. Gao, H. Lin, and J.T Hsieh, [14] presented a new therapeutic approach for Castration-Resistant Prostate Cancer (CRPC) by understanding the onset of Neuro Endocrine Differentiated (NED). Initially, Androgen Deprivation Therapy (ADT) or IL-6 induced lipid accumulation was connected with NED pheno-types. The ADT or IL-6 induces NCD in prostate cancer cells via Adipocyte Differentiation-Related Protein (ADRP), which was a major element of adiposome, and Peroxisome Proliferator-Activated Receptor γ (PPAR γ) was a major feature of lipogenic transcription. Additionally, the

ADRP protein was identified in exosome, which released from the prostate cancer cells. The detected exosomes were attained to induce NED of prostate cancer cells in a paracrine fashion. Understanding the role of ADRP/ PPAR γ in NED provides new targets for CRPC therapy. In this research study, the PPAR γ gene was not identified in both IL-6 treatment and PPAR γ database that affects the post-transcriptional regulation of underlying mechanism.

H.R. Chen, C. DeLisi, Z. Hu, and D.H. Sherr, [15] developed an effective system, which was based on mining a human functional linkage setup for correlating the modules of disease and drug gene targets. The developed architecture comprises of gene mutation, functional connectivity, gene expression, proximity within module gene and multiple information sources. In this research study, the developed system identifies the multi-targeted drug candidates that helps to treat pro-state and breast cancer, and also to correct aberrant cellular functions. The developed system has the ability to deliver a more appropriate drug discovery pipeline. Here, the interaction between the genes occur only in the similar cellular compartment not in others, which was considered as one of the main drawbacks in prostate cancer detection.

R. Li, X. Dong, C. Ma, and L. Liu, [16] developed a new system to identify the prostate related cancer genes in the PPI network using KEGG pathway. The shortest paths between the prostate related cancer genes were found using Dijkstra algorithm. In this literature paper, the developed system delivered an effective performance in identifying the prostate cancer genes in light of elapsed time (second). The experimental outcome confirmed that the surrogate genes were selected to perform an effective classification. In this study, classification corresponds to a minimum characterization of cancer phenotypes on the molecular level. Thus, the experiment results serve as a candidate gene for further functional studies. The shortest paths between the prostate related cancer genes were identified effectively in that a few genes were irrelevant to diseases that need to be concentrated in further research works.

To address the above mentioned issues, a superior approach (modified Ford-Fulkerson algorithm) is developed for determining the shortest path between prostate cancer-related genes.

III. PROPOSED METHODOLOGY

The PPI handles an extensive range of biological practices like developmental control, metabolic and cell to cell interactions. The non-covalent connections between the residue chain is the basis for PPI, protein folding and protein assembly. These non-covalent connections induce a diversity of associations and interactions among the proteins. The PPI is classified in many ways on the basis of their functional properties and contrasting structure. In addition, the PPI is classified as heterooligomeric or homooligomeric on the basis of the interaction surface.

The comprehensive analysis of PPI enables a better understanding of cellular functions, organizations and processes.



A few other applications of PPI network include protein complex identification, interactions reliability estimation, detection of protein interactions, etc. In this research paper, the shortest path between the prostate related cancer genes is identified. This process includes many advantages; earlier detection of prostate cancer, assists the physicians during surgery, cost-efficient compared to the existing machine learning approaches. The step by step procedure of proposed methodology is detailed below and also graphically represented in Fig. 1.

Collect STRING database.

- PPI structure in the STRING database is identified by using KEGG pathway.
- Determine the prostate related genes from the STRING database.
- Identify the shortest path between the prostate cancer-related genes using Modified-Ford-Fulkerson algorithm.

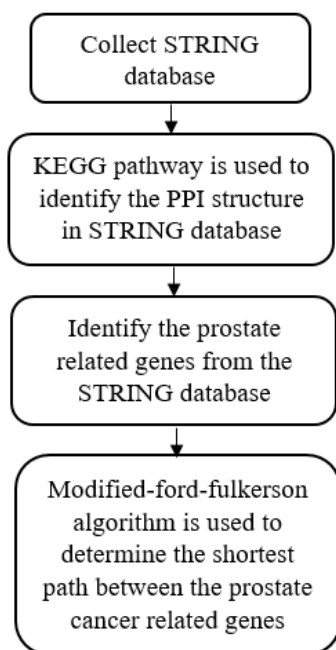


Fig. 1. Work flow of proposed system

A. STRING dataset

The STRING database is a biological database, which comprises of predicted PPI connections. The current version of STRING database (version 10.0) contains information about 9.6 million proteins for 2000+ organisms. In this scenario, the information is collected from numerous resources that include public text collection, experimental information, and computational prediction. The STRING database is a publicly available database, which is utilized to highlight the functional enrichment on the basis of user listed proteins. Here, the enrichment tests are performed by many classification systems, a few classification systems are listed below,

- KEGG pathway
- Gene Ontology
- InterPro
- Pfam, etc.

In STRING database, the weighted graph is plotted on the basis of gene weight values that are graphically denoted in

Fig. 2. The PPI network is the essential component for determining the shortest path between the disease related genes. The PPI network is used to filter the active genomics genes that provides platform for annotating functional, structural and evolutionary characteristics of the proteins. Investigating the connection of PPI provides a new direction for future experimental research and also it gives the cross species for effective interaction mapping. Also, the STRING database stores the predicted PPI from interaction transferred from model organisms based on orthology, text mining of scientific texts, and interactions computed from genomic features. In this experimental research, KEGG pathway is suggested as a classification system, because it effectively refers to the PPI structures.

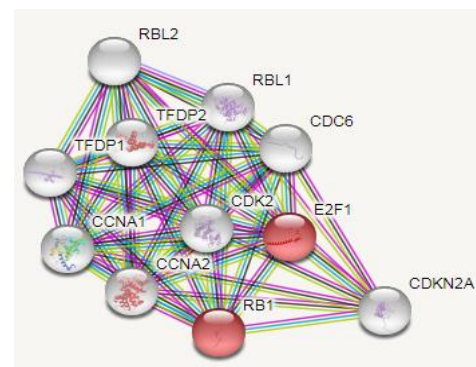


Fig. 2. Sample PPI connection for E2F1gene

B. KEGG path enrichment analysis

The KEGG pathway is a database reservoir that comprises of systemic functional, genomic, and chemical information. Particularly, the gene catalogs from sequenced genomes are combined with higher level systemic functions of the ecosystem, the cell and the organism. Several efforts are done to develop a knowledge-based system for systemic functions by organizing and capturing the experimental knowledge in computable forms, which is named as KEGG modules, BRITE functional hierarchies, and KEGG pathway maps. Various efforts are done for improving and developing the cross-species annotation process to link genomes to the networks of molecules using KEGG orthology system.

As a result, KEGG pathway is extensively utilized as a reference knowledge base to interpret and integrate large scale databases for generating high throughput experimental techniques and the sequences of genome. The KEGG pathway is expended to more practical applications like health based substances, drugs and integrating human disease for maintaining the aspects to support basic research.

C. Prostate cancer-related genes

Initially, the prostate related eighteen genes (BCL2, CCND1, CCNE1, CCNE2, CDKN1A, CDKN1B, CHUK, CREBBP, E2F1, IKBKB, IKBKG, MDM2, NFKB1, NFKBIA, PTEN, RB1, RELA, and TP53) are compiled from the STRING dataset using KEGG pathway, which is detailed in Table 1. The weight of PPI network is constructed on the basis of data from STRING (version 10.0) database.

Detection of Prostate Cancer Related Genes using Modified Ford-Fulkerson Algorithm in Protein-to-Protein Interaction Network

Each and every interaction in the STRING database is determined by an interaction confidence score, which ranges from 1 to 999 for quantifying the likelihood that an interaction may occur. In this research study, Modified-Ford-Fulkerson algorithm is used to determine the shortest path between each protein pairs corresponding to eighteen genes in the PPI network. At last, all the proteins present on the shortest paths are ranked according to the betweenness score.

Table I. Prostate related genes

Pathway description	Count	False discovery rate	Gene ID
Prostate cancer	18	7.70E-31	BCL2, CCND1, CCNE1, CCNE2, CDKN1A, CDKN1B, CHUK, CREBBP, E2F1, IKBKB, IKBKG, MDM2, NFKB1, NFKBIA, PTEN, RB1, RELA, and TP53.

D. Ford-Fulkerson algorithm

The Ford Fulkerson algorithm is a greedy algorithm that determines the maximum flow in a network. In this experimental research, the Ford Fulkerson algorithm is utilized to determine the shortest path between the prostate cancer-related genes, either it comprises of negative or positive value. The single cycle of the Ford Fulkerson algorithm determines the shortest path in the vertices. The Ford Fulkerson algorithm is used to the directed graph of the PPI network, which is constructed for the data obtained from the STRING database. Let be the capacity and for each edge from p to q , let $f(p, q)$ be the flow and $G(V, E)$ be a graph. Then, identify the maximum flow from the sink to the source s . The following procedures need to flow for each and every step in the Ford Fulkerson algorithm.

Capacity constraints: The flow of the edge can-not exceed its capacity that is given in the Eq. (1)

$$\forall (p, q) \in E f(p, q) \leq c(p, q) \quad (1)$$

Skew symmetry: The net flow from p to q is opposite to the net flow from q and p , which is mathematically given in the Eq. (2).

$$\forall (p, q) \in E f(p, q) = -f(q, p) \quad (2)$$

Flow conservation: The net flow to a node is zero, if is s or. The flow conservation is statistically represented in the Eq. (3).

$$\forall_{p \in V} : p \neq s \text{ and } p \neq t \Rightarrow \sum_{w \in V} f(p, w) = 0 \quad (3)$$

After every round in the Ford Fulkerson algorithm, the flow of the network is a legal. Here, the residual network $G_f(V, E_f)$ is a network with capacity $c_f(p, q) = c(p, q) - f(p, q)$, which has no flow. It occurs only when the flow from p to q is in the residual network.

Otherwise,

If $f(p, q) > 0$ and $c(v, u) = 0$, then the residual network capacity $c_f(p, q)$ is rewritten as shown in the Eq. (4).

$$c_f(p, q) = c(p, q) - f(p, q) = f(p, q) > 0 \quad (4)$$

1) Pseudo code of Ford Fulkerson algorithm

Input: Graph G , flow capacity, sink node t and source node s .

Output: Flow f from s to t is a maximum one,

$$f(p, q) \leftarrow 0 \text{ for all edge } (p, q)$$

When there is a path P from s to t in G_f , such that $c_f(p, q) > 0$ for all edges $c_f(p, q) > 0$.

For each edge $(p, q) \in P$

$$\text{Find } c_f(P) = \min\{c_f(p, q) : (p, q) \in P\}$$

For each edge

$$f(p, q) \leftarrow f(p, q) + c_f(P) ; \text{ flow along the path}$$

$$f(p, q) \leftarrow f(p, q) - c_f(P) ; \text{ flow might be return later}$$

If S is the set of nodes reachable in the residual network then the total flow found from s to t is equal to the total capacity in the original network of edges from s to the remainder of V . If the graph $G(V, E)$ has multi sinks and sources, it is expressed as given in the Eq. (5) and (6).

$$T = \{t \mid t \text{ is a sink}\} \quad (5)$$

$$S = \{s \mid s \text{ is a sink}\} \quad (6)$$

Now, add a new source s^* with an edge (s^*, s) from s^* to each node $s \in S$, then the capacity constraint is updated as shown in the Eq. (7). Respectively, add a new sink t^* with an edge (t^*, t) from t^* to each node $t \in T$, then the capacity constraint is updated as presented in the Eq. (8).

$$f(s^*, s) = d_s \left(d_s = \sum_{(s, p) \in E} f(s, p) \right) \quad (7)$$

$$f(t^*, t) = d_t \left(d_t = \sum_{(q, t) \in E} f(q, t) \right) \quad (8)$$

Finally, if each node p has constraint d_p , replace the nodes p_{in} and p_{out} with a capacity $f(p_{in}, p_{out}) = d_p$.

E. Modified Ford Fulkerson algorithm

The modified Ford Fulkerson algorithm computes the maximum flow in a network and also it analysis the all possible paths between the given pair of vertices (genes). The Ford Modified Fulkerson algorithm reconstructs the path with the small modification, a new iterative matrix is utilized in the Ford Fulkerson algorithm for removing the invalid paths.



- No direct acknowledgement path (weight of the path is not equal to the reverse path).
- No reverse path between the two vertices, so there is no direct path existing between the vertices.

The pseudo code of modified Ford Fulkerson algorithm is represented below,

1) Pseudo code of modified Ford Fulkerson algorithm

Input: Graph G , flow capacity c , sink node and source node s .

Output: Flow from $stot$ is a maximum one,

$$f(p, q) \leftarrow 0 \text{ for all edge } (p, q)$$

When there is a path P from $stot$ in G_f such that

$$C_f(p, q) > 0 \text{ for all edges } (p, q) \in P.$$

For each edge $(p, q) \in P$

$$\text{Find } c_f(P) = \min\{c_f(p, q) : (p, q) \in P\}$$

$$f(p, q) \leftarrow f(p, q) + c_f(P); \text{ flow along the path}$$

$$f(p, q) \leftarrow f(p, q) - c_f(P); \text{ flow might be return late}$$

IV. EXPERIMENTAL RESULT AND DISCUSSION

The proposed methodology was simulated using MATLAB (version 2017a) with 3.0 GHz Intel i5 processor, 1TB hard disc, and 8 GB RAM. In this research study, the prostate related genes were undertaken for experimental analysis such as, BCL2, CCND1, CCNE1, CCNE2, CDKN1A, CDKN1B, CHUK, CREBBP, E2F1, IKBKB, IKBKG, MDM2, NFKB1, NFKBIA, PTEN, RB1, RELA, and TP53. The proposed methodology performance was compared with other shortest path identification approaches like Dijkstra algorithm [16] and Ford Fulkerson algorithm on a reputed dataset: STRING for evaluating the efficiency of proposed technique. The proposed methodology performance was evaluated in light of elapsed time.

A. Prostate cancer related genes

Usually, the prostate cancer occurs in the small walnut shaped gland in men that produces seminal fluid, which nourishes and transports sperms. The prostate cancer is the second leading cancer in men. In this scenario, the prostate related eighteen genes (BCL2, CCND1, CCNE1, CCNE2, CDKN1A, CDKN1B, CHUK, CREBBP, E2F1, IKBKB, IKBKG, MDM2, NFKB1, NFKBIA, PTEN, RB1, RELA, and TP53) are collected from the SRING dataset using KEGG pathway. For instance, the gene structures of BCL2, IKBKG, and CCND1 are graphically denoted in the Fig. 3, 4, and 5.

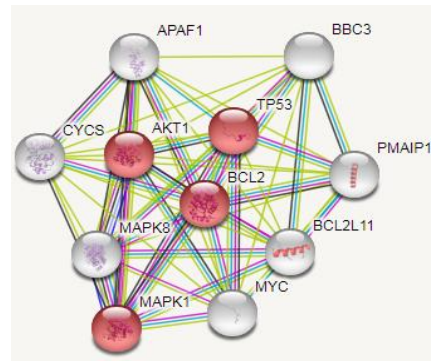


Fig. 3. Gene representation of BCL2

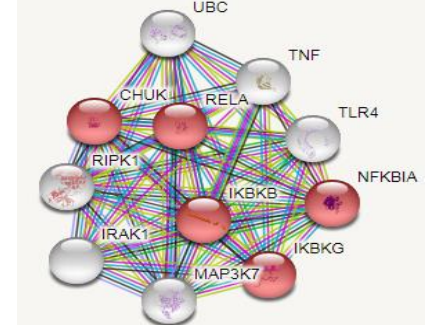


Fig. 4. Gene representation of IKBKG

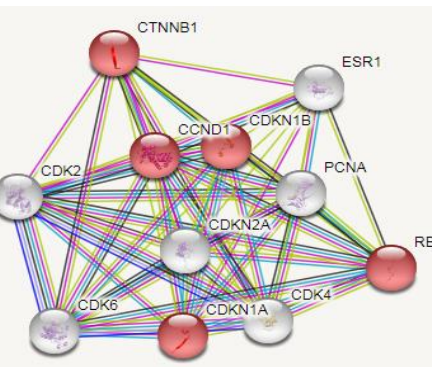


Fig. 5. Gene representation of CCND1

B. Quantitative analysis

Performance measure is defined as the regular measurement of results and outcomes, which develops a reliable information about the effectiveness of the proposed methodology. Table 2 describes the performance analysis of the proposed (modified Ford-Fulkerson algorithm) and existing technique (Ford-Fulkerson algorithm) by means of elapsed time. The modified Ford-Fulkerson algorithm is faster in identifying the shortest path between the prostate cancer related genes compared to Ford-Fulkerson algorithm, because it eliminates the unwanted paths like no direct acknowledgement path and no reverse path between the two vertices. The graphical depiction of the existing and proposed methodology is given in Fig. 6.

The collected prostate related genes from STRING dataset is given as the input for mat-lab. Normally, there are two stages involved in developing a routing table for link-state routing protocols.



Detection of Prostate Cancer Related Genes using Modified Ford-Fulkerson Algorithm in Protein-to-Protein Interaction Network

In first stage, the whole network map is stored in each router and in second stage, the shortest distance of each node (gene) is calculated by using modified Ford-Fulkerson algorithm. Fig. 6 represents the graphical depiction of existing and proposed approach in terms of “genes vs elapsed time”. Here, the gene ranges from 1 to 18, so 1 is BCL2 gene, 2 is CCND1 gene, 3 is CCNE1 gene, 4 is CCNE2 gene, 5 is CDKNIA gene, 6 is CDKN1B gene, 7 is CHUK gene, 8 is CREBBP gene, 9 is E2F1 gene, 10 is IKBKB gene, 11 is IKBKG gene, 12 is MDM2 gene, 13 is NFKB1 gene, 14 is NFKBIA gene, 15 is PTEN gene, 16 is RB1 gene, 17 is RELA gene, and 18 is TP53 gene.

Table II. Performance evaluation of proposed and existing approach in light of elapsed time

Gene type	Elapsed time (second)	
	Ford-Fulkerson algorithm	Modified ford-Fulkerson algorithm
BCL2	0.004454	0.004381
CCND1	0.002206	0.001940
CCNE1	0.003127	0.002999
CCNE2	0.008706	0.007498
CDKNIA	0.004405	0.004313
CDKN1B	0.011109	0.010357
CHUK	0.003165	0.002756
CREBBP	0.011109	0.010779
E2F1	0.009446	0.008930
IKBKB	0.008310	0.007518
IKBKG	0.003152	0.003120
MDM2	0.004518	0.004135
NFKB1	0.003177	0.002985
NFKBIA	0.003177	0.003150
PTEN	0.004432	0.004383
RB1	0.001999	0.001915
RELA	0.003140	0.003130
TP53	0.023240	0.010156

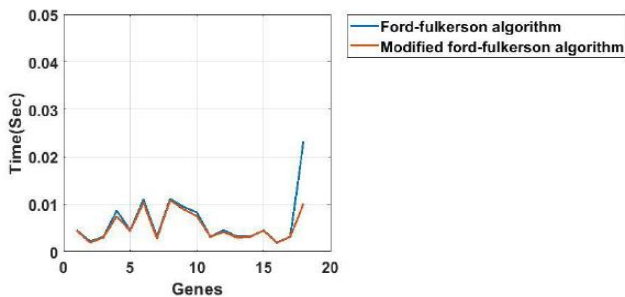


Fig. 6. Graphical comparisons of existing and proposed methodologies

Compared to existing method (Ford-Fulkerson between the prostate related genes. The modified Ford-Fulkerson algorithm identifies a shortest path among all pairs of genes (nodes) that does not contains any cycles of negative length. The graphical representation of BCL2 gene, IKBKG gene, and CCND1 gene shortest path for prostate cancer are shown in Fig. 7, 8 and 9.

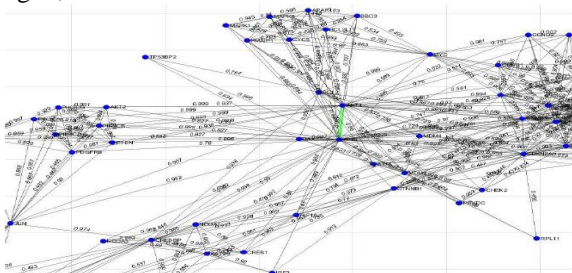


Fig. 7. Shortest path representation of BCL2 gene

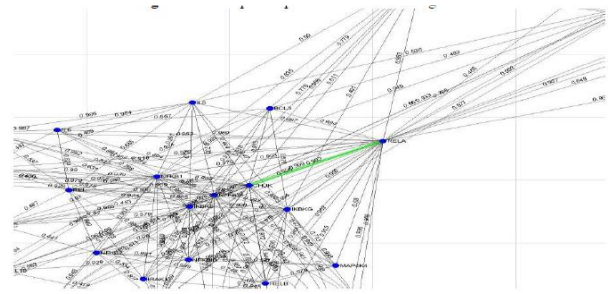


Fig. 8. Shortest path representation of IKBKG

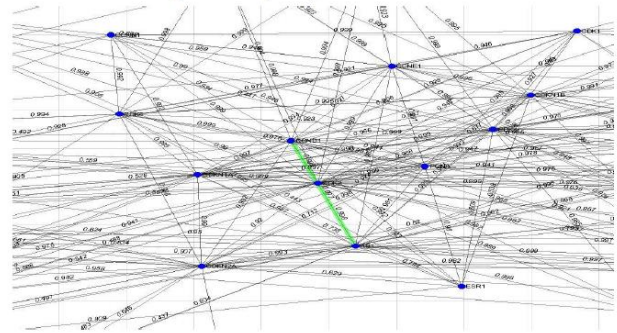


Fig. 9. Shortest path representation of CCND1 gene

Table 3 represents the investigation about existing and proposed methodology. R. Li, X. Dong, C. Ma, and L. Liu, [16] developed an effective methodology (Dijkstra’s algorithm) for identifying the shortest path of prostate related genes in PPI network. A weighted PPI network was constructed with 18 prostate related genes based on the PPI data from STRING database, which was retrieved using KEGG pathways. This experiment was performed on an online available database (i.e., STRING database) for validating the result in light of elapsed time. For instance, the existing work averagely consumed 0.0294 second for finding the shortest path between 18 genes. Whereas, the proposed work consumed only 0.0052 second that was lower than the existing work (Dijkstra’s algorithm). Table 3 confirmed that the proposed work performed effectively than the existing shortest path approach, which is graphically denoted in Fig. 10.

Table III. Analysis of proposed and existing approach

Gene type	Elapsed time (seconds)	
	Dijkstra algorithm [16]	Modified Ford-Fulkerson algorithm
CL2	0.027302	0.004381
CCND1	0.026307	0.001940
CCNE1	0.027406	0.002999
CCNE2	0.031214	0.007498
CDKNIA	0.030343	0.004313
CDKN1B	0.029897	0.010357
CHUK	0.029709	0.002756
CREBBP	0.029637	0.010779
E2F1	0.029005	0.008930
IKBKB	0.030037	0.007518
IKBKG	0.0298667	0.003120
MDM2	0.029735	0.004135
NFKB1	0.029945	0.002985
NFKBIA	0.029866	0.003150
PTEN	0.029529	0.004383
RB1	0.030195	0.001915

RELA	0.02997	0.003130
TP53	0.029742	0.010156
Average	0.0294	0.0052

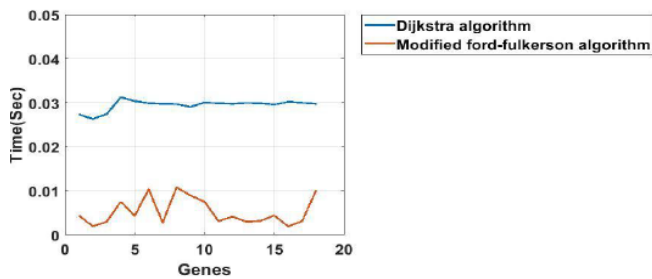


Fig. 10. Analysis of existing and proposed methodology

V. CONCLUSION

In this research study, a new computational approach was proposed on the basis of a PPI network to find the shortest path between the prostate cancer-related genes. Here, the proposed approach was applied to prostate cancer for finding the shortest path between eighteen prostate cancer genes in the PPI network, which was constructed based on STRING data. Analysis of these shortest path genes comprises of numerous advantages; assists the clinicians during surgery, earlier diagnosis of prostate cancer, and cost-efficient associated with other existing machine learning methodologies. Related to other existing approaches in prostate cancer detection, the proposed approach delivered a superior performance by means of elapsed time. The proposed system reduced the elapsed time up to 0.025-0.002 (seconds) compared to the existing methods. In future work, an effective distance-based approach was developed for further reducing the elapsed time to find the shortest path between the prostate cancer-related genes.

REFERENCES

1. F. Yuan, Y. Zhou, M. Wang, J. Yang, K. Wu, C. Lu, X. Kong, and Y. D. Cai. (2015). Identifying new candidate genes and chemicals related to prostate cancer using a hybrid network and shortest path approach. *Computational and mathematical methods in medicine*.
2. N. Puthiyedth, C. Riveros, R. Berretta, and P. Moscato. (2015). A new combinatorial optimization approach for integrated feature selection using different datasets: A prostate cancer transcriptomic study. *PLoS one*, 10(6), pp. e0127702.
3. O. E. Bryzgunova, M. Y. Konoshenko, and P. P. Laktionov. (2018). MicroRNA-guided gene expression in prostate cancer: Literature and database overview. *Journal of gene medicine*, 20(5), pp. e3016.
4. R. A. Eeles, Z. Kote-Jarai, G. G. Giles, A. A. Al Olama, M. Guy, S. K. Jugurnauth, S. Mulholland, D. A. Leongamornlert, S. M. Edwards, J. Morrison, and H. I. Field. (2008). Multiple newly identified loci associated with prostate cancer susceptibility. *Nature genetics*, 40(3), pp. 316.
5. R. Zhao, Y. Wang, M. Zhang, X. Gu, W. Wang, J. Tan, X. Wei, and N. Jin. (2017). Screening of potential therapy targets for prostate cancer using integrated analysis of two gene expression profiles. *Oncology letters*, 14(5), pp. 5361-5369.
6. C. Yajun, T. Yuan, W. Zhong, and X. Bin. (2018). Investigation of the molecular mechanisms underlying postoperative recurrence in prostate cancer by gene expression profiling. *Experimental and therapeutic medicine*, 15(1), pp. 761-768.
7. M. C. Albertini, F. Olivieri, R. Lazzarini, F. Pilolli, F. Galli, G. Spada, A. Accorsi, M. R. Rippo, and A. D. Procopio. (2011). Predicting microRNA modulation in human prostate cancer using a simple String Identifier (SID1. 0). *Journal of biomedical informatics*, 44(4), pp. 615-620.
8. L. Gong, D. Zhang, Y. Dong, Y. Lei, Y. Qian, X. Tan, S. Han, and J. Wang. (2018). Integrated Bioinformatical Analysis for Identifying the

9. D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K. P. Tsafou, and M. Kuhn. (2014). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic acids research*, 43(1), pp. D447-D452.
10. Q. C. Zhang, D. Petrey, J. I. Garzon, L. Deng, and B. Honig. (2012). PrePPI: a structure-informed database of protein-protein interactions. *Nucleic acids research*, 41(D1), pp. D828-D833.
11. S. Detchokul, E. J. Crampin, A. Elangovan, A. G. Frauman, and M. J. Davis. (2015). Network analysis of an in vitro model of androgen-resistance in prostate cancer. *BMC cancer*, 15(1), pp. 883.
12. L. Chen, Z. Zhang, Y. H. Zhang, Y. D. Cai, and T. Huang. (2018). Inferring novel tumor suppressor genes with a protein-protein interaction network and network diffusion algorithms. *Molecular Therapy-Methods & Clinical Development*, 10, pp. 57-67.
13. Y. Lin, L. Shen, F. Chen, X. Tang, Z. Sun, C. Du, J. Chen, H. Ding, and B. Shen. (2018). Biomarker microRNAs for prostate cancer metastasis: screened with a network vulnerability analysis model. *Journal of translational medicine*, 16(1), pp. 134.
14. L. C. Lin, C. H. Lai, A. C. Gao, H. Lin, and J. T. Hsieh. (2017). Induction of neuroendocrine differentiation in castration resistant prostate cancer cells by adipocyte differentiation-related protein (ADRP) delivered by exosomes. *Cancer letters*, 391, pp. 74-82.
15. H. R. Chen, C. DeLisi, Z. Hu, and D. H. Sherr. (2016). A network based approach to drug repositioning identifies plausible candidates for breast cancer and prostate cancer. *BMC medical genomics*, 9(1), pp. 51.
16. R. Li, X. Dong, C. Ma, and L. Liu. (2014). Computational identification of surrogate genes for prostate cancer phases using machine learning and molecular network analysis. *Theoretical Biology and Medical Modelling*, 11(1), pp. 37.

AUTHORS PROFILE



Mr. SANJEEV PRAKASHRAO KAULGUD, Currently pursuing Ph.D. on Data mining and Bio-informatics at REVA University, Bengaluru. Working as Assistant Professor in PRESIDENCY University, Bengaluru. Having 8 years of experience in teaching profession, Industry and research. His Research is based on detecting Cancer in the early stages using Protein Sequences of humans.



Dr. VISHWANATH HULIPALLED, presently working as Deputy Dean, Research Department, REVA University. He has received his Ph. D. degree from Jawaharlal Nehru Technological University, Hyderabad, India. He has a teaching experience of more than 19 years. His research areas are Time Series Mining, Big Data Analytics, Cloud Computing and more. Member for both Board of Studies and Board of Examiners since 2008 in University Visvesvaraya College of Engineering, Bangalore.



Prof. S. S. PATIL, Professor of Computer Science, Head and University Head, Dept. of Agril Statistics, Applied Mathematics & Computer Science, University of Agricultural Sciences, Bangalore. He obtained his Doctoral degree in Computer Science. He is currently Professor and leading a research team in the Dept. of Computer Science. He has a teaching experience of more than 25 years. Member for Board of Studies and Board of Examiners of several Universities.

