Quality of Service Enhanced Framework for Disease Detection and Drug Discovery

A. Singh[1*], S. Sharma[2], R. Singh[3], G. Singh[4], A. Kaur[5]

[1]Dept. of Computer Science, Guru Nanak Dev University, Amritsar, India
[2] Dept. of Computer Science, Hindu College, Guru Nanak Dev University, Amritsar, India
[3,4]Dept. of Computer Science, Guru Nanak Dev University, Amritsar, India
[5] Dept. of CSE, DBU, Mandi Gobindgarh, India

[*]*Corresponding Author: amritsinghmand@gmail.com, Tel.: +91-9988080241*

Abstract— Disease detection frameworks having blend of distinct fields of concerned research are the need of hour as the domain is very challenging and essential for the human race. Its expanding every day and its impact is much bigger than the earlier era. Quality of service(QoS) is an important consideration for designing sophisticated futuristic healthcare systems. One such quality parameter of reliability is consolidated in this disease detection framework incorporating information-centric networking of medical knowledge-base and protein/gene knowledge-base. The proposed framework is designed and implemented keeping in view of research progress of two highly interrelated research areas which are often not used together but can be combined to formulate an efficient disease detection system. Machine learning and deep learning is also incorporated to improve the quality of service parameter in terms of accuracy and reliability of the system. Various ML configurations formed a test bed for the accuracy check at different system settings indicating a perfect mix for a specific detection requirement. Results indicate promising detection rate of 71% combined accuracy and a remarkable 97% accuracy rate for specific single class detection along with the improvement in reliability factor.

Keywords— ICN, Disease detection, Drug discovery, Machine Learning, Quality of Service.

I. INTRODUCTION

Health care systems are developed and worked upon for excellence since ages. Emergence of new technology boosts the process of up-gradation and refinement of such systems. This research focus on exploring some of the avenues of advanced healthcare system setups.

First focus is integration of genetic biomarkers of disease detection in the form of use of existing and evolving research with conventional disease detection strategies and drug discovery process.

Second concern is use of more suitable ICN for exchange of information between various health care modules. Third important consideration is enhancing QoS parameters by exploring previous benchmarks for their QoS parameters limitations and incorporating their remedies in ICN layer itself.

Fourth and final task is design of a framework which integrates the aforementioned tasks and fulfils important QoS parameters of reliability and scalability among other parameters.

II. RELATED WORK

Bengt Ahlgrem. et al. [1] in 2012 conducted a review of information centric networking (QoS) thereby elucidating two important factors viz. Scalability and cost of content distribution. The main development to introduce scalability and limit cost of communication is adoption of content distribution networks (CDN). Authors have surveyed network approaches that introduced hop to hop transport along with caching and enhanced reliability of network.

Oscar Gama et al. in 2008 revealed the need of providing Quality of Service for e-healthcare systems. They have surveyed several mechanisms using wireless sensor networks to improve certain QoS parameters like packet loss ratio, transmission delay, bandwidth availability and minimum sampling rate for ECG signals.

In 2012 A. A. Hamed et al. [6] proposed T-Recs System which was the very interesting concept of drug recommender system for users. It was time aware twitter-based drug recommender system which uses the tweet sentiments at regular time stamp and based on that tweets data classifier

provide the recommendations using white box approach decision tree approach. Using the data mining approach, the system provided the alternate medicine product to the user. Their recommendations using this approach promoted public health as well as awareness in the public domain.

In 2012 R.C. Chen at al. [7] proposed the recommender system for anti diabetic drugs selection based on domain ontology and the SWRL (sementic web rule language). In proposed recommender system the medicine consultant data/information was transferred in to the anti diabetic drugs knowledge system. The system has the capability to scrutinize the symptoms of diabetetes and recommend the appropriate drug to user among the related drugs. The medication recommender system, based on domain ontology, employ the knowledge and facilitate the user as well as helps in practice for the management of diabetes. The recommender system acquired the help of SWRL (semantic web rule language) and the JESS (Java expert system shell) to prescribe the prescription for the patients.

In 2012 R. Chen et al. [8] expressed the need of personalised medicines and suggested the system biology which describe the personalized medicines for future health care system on individual patients. The author suggested that the advancement in the technology boosted the area's of personalised medicine, which shifted the paradigms of health care systems from disease detection and its treatments to the predictive as well as preventive medicines also the personalized health recommender system. The author also envisioned that the whole genome would be the colaborative part of the patient medical record system and the health of the patient regularly examined at home or at clinic and the data generated through regular monitering system will be stored in DBMS's at ending station like hospitals, and the medical experts at hospitals can use it for individual diagnosis and treatment purposes as well as can act as the health information service provider.

In 2012 L.M. Ma'aruf at al. [9] proposed the expert diet prescription system which identified the illness by directly or through symptoms and recommend the appropriated diet or diet plan corresponding to the illness. The author enlighted the use of fruits/food instead of Drugs because drugs have to much side effects of the human body, so they recommended the need of diet expert system instead of drug recommender system because the long queues at hospitals have alarming situation. The author considered the various illness and creates the database or knowledge base for malaria, diabetes, goiter, cancer, measles, heart disease and cholera, and developed the recommender system with the help of MYSQL, PHP and wamp server. Their expert diet system can be used by anyone at anywhere either for professional use or even for personal use. The system have great benefit towards the society, and improve the lifestyle of people those who properly utilized it.

In 2012 C. Doulaverakis et al. [10] proposed the drug recommender system based on ontology. They suggested that how sementic web technology and OWL is a better match for drug recommender system. In the ontology specially OWL can easily encapsulate or express the medical issues or the medical information and the rule based reasoning can easily encode as well as capture the drug-drug interation information which can be helpful for medical decisions as well as assistance to expert of domain. The author used the GalenOWL the sementic web based technology for their discovery. They express their views over the importance to use semenctic web and OWL for drug recommender server. They qualitatively compare and comparitatively express the GalenOWL and GalenD rules, their prediction about sementic web technology as compare to traditionally business logic revealed the importance of sementic web on the basis of time and memory requirements. Finaly they revealed that Galen OWL is also capable for drug-disease interaction discovery.

In 2012 A. S. Hussein et al. [11] proposed the reliable and accurate recommender system for the CDD (chronic disease diagnosis) specially in the case of diabetes. They use the decision tree classification approach to grab the peak accuraccy in disease risk prediction and random forest algorithm for mining their recommendation. To provide an accurate recommendation on medical data is a challenging task because of the real-time complex, large, unbalence, missing and noisy data. So they consider the attribute selection methodology to reduce the complexity of real-time unbalence data which leads improvement in prediction of the recommended model. They suggested how information and communation technology growth and its advancement opens the doors for researcher in the areas of E-Health monitering systems. Those who used the emerged recommender system, their lifestyle and health got improved because of its predicting capability and appropriate recommendations for consumers.

In 2005 K. Shimadaa et al. [12] proposed the drug recommender system for patients those who have infectious diseases. They proposed the decision support system which helps the doctors to select first line drug in appropriate way. Their recommender system classifies the patient capability to resist themselves from transferable/infectious diseases by considering risk level for infection.

In 2014 Y. Zhangy et al. [13] proposed the COMER (cloud based medicine recommendation) which is cloud-based recommender system for online purchasing the medicines. The author expressed their views on QoE (Quality of Experience) and suggested the importance of cloud-based recommender system in case of personalized medicine recommendations. With the rapid growth in e-commerce people prefer to buy medicines online because of the easiness of such broad area, but to buy medicine online blindly having

a serious issue without any medication guidance because of the ineffectiveness and reliability as well as its validity. So, author expressed the need of reliable medication system which should recommend cloud- based medicine recommendations and enrich the user with QoE (Quality of Experience). The author used the collaborative filtering and tensor decomposition as well as clustering approach for their recommendation which recommend the user with topmost related medicines related with their symptoms.

In 2015 B. Yoosooka et al. [14] proposed the medication recommender system based on ontology which is a semantic web technology. They suggested that how sementic web technology and OWL is a better match for drug recommender system. The author expressed the need of medical recommender system because non priscription medication could lead some adverse effects when consumed so author proposed the medication recommender system. The author considered the ontology, PHP with MySql database for medication recommendation which is workable on the PC as well as on mobile suitable environment and achieve the user satisfaction approximate of 90% with standard deviation of 0.51 which is highly statisfiable.

Table 1. Comparison of Benchmarks with respect to QoS parameters, **TD**: Transmission Delay, **B**: Bandwidth, **PD**: Packet, **SL**: Service Level, **MQ**: Monitoring Quality, **Rel.:** Reliability, **AL**: Application Lifetime, **PL:** Packet Loss, **AD:** Acceptable Delay, **PC:** Power Consumption

Features / Bench-marks	Quality of Service features/parameters incorporated									
	TD	*B*	*PD*	*SL*	*MQ*	*Rel.*	*AL*	*PL*	*AD*	*PC*
H.She et al. [1]	√	√	×	×	×	×	×	×	×	√
Vergados et al. [2]	×	√	√	√	×	√	×	×	×	×
Gondal et al. [3]	×	×	×	×	√	√	×	×	×	×
Milan et al. [4]	×	×	×	×	√	×	√	×	×	×
Henrion et al. [5]	×	×	×	×	×	×	×	√	√	×
Hamed et al. [6]	×	×	×	×	×	√	×	×	×	×
Chen et al. [7]	×	√	√	√	×	√	√	×	×	×
Rui Chen et al. [8]	×	×	×	×	√	×	×	×	×	×
Maaruf et al. [9]	√	√	√	√	×	√	√	√	×	√
Doulaverakis et al. [10]	√	√	√	×	×	√	√	×	√	×
Hussein et al. [11]	×	×	×	×	√	√	√	√	×	×
Shimda et al. [12]	×	×	×	×	×	×	×	√	×	×
Zhangy et al. [13]	√	√	√	√	√	√	×	×	√	×
Yoosooka et al. [14]	√	√	√	×	×	√	√	×	×	×
Janga et al. [15]	×	√	√	√	×	√	×	×	×	×

In 2014 S. C. Janga et al. [15] proposed the system for personalised medicine based on netwrok approahes which plays very important role in medicine as well as in other biotechnology areas. Network interaction between diverse cellular entities like metabolites, RNA, DNA and Proteins are very complex but with advancement in the technology boosted the current knowledge to understand such complex structure and their interaction, which leads proper understanding of their functionalities in living organism. This understanding further helps in disease detection at a level of cellular entities through netwrok based framework. Thus this multi scaled personalized medication approaches guides the upcoming generation to understand the basic functionalities of living organism.

Resource utilization can be done in a better way while designing a system in a cloud environment with the help of virtulization. Quality of service can be enhanced by considering threats and removing them at the initial designing of the system [16].

Multiobjective optimization can be helpful in designing a detection system keeping in view all the contributing factors and leading to fulfill multiple objectives [17].

The summary of recent literature along with the respective QoS features achieved is represented in Table 1.

III. METHODOLOGY

Collection of information from various data sources and integrating them in QoS enhanced ICN (Information Centric Network) module.
ICN stack can be used to integrate Body Sensor Network (BSN) information first up. This enhances QoS parameter of reliability

One part of information is passed to the domain expert module. This part of information contains Medical Knowledgebase Test report, BSN inputs. This process helps in providing accurate information to the domain expert and eliminate majority of parameters which may degrade QoS.

Diagnosis obtained from domain expert is feed to disease detection and recommender system. This system compares the diagnosis of doctor/domain expert with its findings. This module work on information from QoS enhanced ICN module. Which in turn receive information about genes responsible for a disease from protein class/gene prediction module. Protein /gene database gets data via web scrapping tools by exploration of websites like NCBI, UNIPROT etc.

Protein class/Gene prediction module predicts the protein class from the user protein tests obtained by hydrolysis

process. This indicates the presence or absence of disease causing gene in the patient's body.

Drug discovery module get information from medicine database via QoS enhanced ICN module and find out suitable drug for the cure.
Thus, overall framework gets information passed from QoS parameters hence its more reliable making the entire system reliable.

Final learning module employs deep learning in order to continually improve the system by getting information about QoS and their effect on the framework and can decide on enforcing a parameter mix for better result.
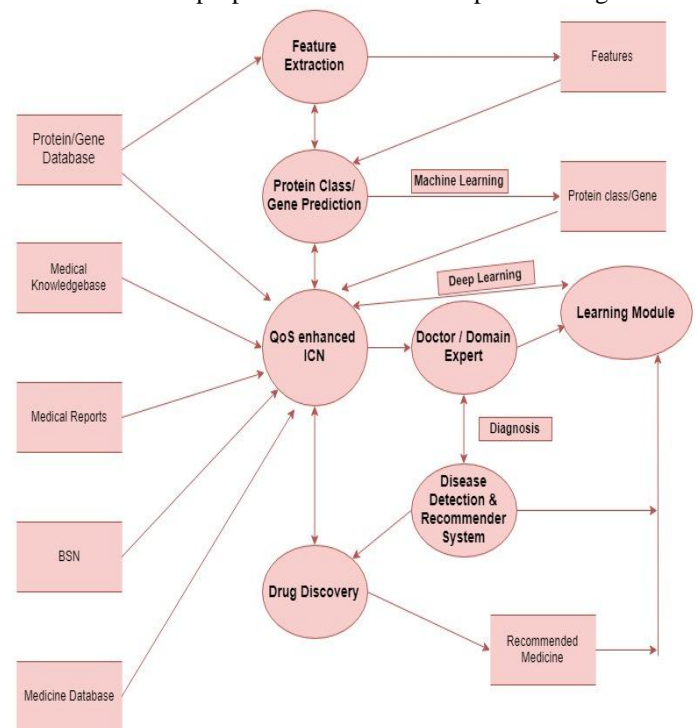Workflow of the proposed framework is depicted in Figure 1.

Figure 1. Proposed Framework for Quality of service enhancement in medical information-centric networking for disease detection recommender system.

IV. RESULTS AND DISCUSSION

Results obtained clearly indicate the usability of the system framework:

Amino acid sequence can be extracted from hydrolysis of protein samples obtained from patient then protein class can be predicted using SDFs. Use of this approach make genetic tests an affordable option in DDS and is one of the important contribution of this research.

Gene identification system works by web scrapping of various web resources containing information about genes responsible for various diseases. This is linked with Machine learning for the update of database as new information about genes is available on the web resources, for enhancing the use of the framework. UNIPROT standard annotated and verified gene data for various diseases and then for various protein classes were taken for initial setup to make the system reliable.

Another major contribution of this research is identification of a critical limitation of Machine Learning approach in terms of biasness in results. The solution to this limitation was devised and implemented. Protein class detection was performed with various settings and best results were obtained by a configuration of 12 classes respectively for 70 sequences using 34 features using random forest algorithms with an overall accuracy rate of 71.09% shown in Table 1 and devising a mechanism of machine learning configurations to push the accuracy rates to a remarkable 97% (average) individual class detection, depicted with the confusion matrix for run result of WEKA shown in Figure 2.

Table 1. Protein Classes prediction accuracy with different settings

ML Classifiers	Feature and sequence settings for 12 Protein Classes		
	20 Features, 70 Sequences	24,70	34,70
J48 decision tree	42.92	45.35	50.13
Random Forest	58.96	59.10	71.09

```
=== Confusion Matrix ===

a b c d e f g h i j k l  <-- classified as
5 0 1 0 0 0 0 0 0 0 0 0 | a = Defensin
0 4 0 0 1 0 0 0 1 0 0 0 | b = AcidPhosphatase
0 0 3 0 0 0 2 0 1 0 0 0 | c = VoltageGatedChannel
0 0 0 1 1 2 0 1 0 1 0 0 | d = DNARepairProtein
0 0 0 0 6 0 0 0 0 0 0 0 | e = Decarboxylase
0 0 0 0 0 4 0 1 0 0 0 0 | f = HeatShockProtein
0 0 1 0 0 0 5 0 0 0 0 0 | g = Aminopeptidase
0 0 0 0 1 1 0 3 0 1 0 0 | h = G-Protein
0 0 0 0 0 0 0 0 6 0 0 0 | i = WaterChannel
0 1 0 1 1 0 0 1 0 2 0 0 | j = NucleotidylTransferase
0 0 0 0 0 0 0 0 0 0 4 1 | k = BCellAntigenReceptor
1 0 0 1 0 0 0 0 0 1 1 2 | l = CellSurfaceReceptor

=== Run information ===
```

Figure 2. Screenshot of confusion matrix obtained for individual class prediction using WEKA

Gene matching was implemented using simple search algorithm, searching for disease related genes in the predicted protein class from the patient sample. Further, ANN was used to speed up the matching process by anticipating various mix and matches, increasing the search efficiency suitable for bigger systems to be incorporated by this framework.

Recommender system creates ranking of the matched genes and gives correctness score of disease detection. For the higher match percentage higher score is generated. A deep learning approach integrate it with learning-module/follow-up system verifying detections with actual results and make the system capable of self-learning. Figure 3 show the participation of each protein class in causing cancer.

The proposed system plays a support role to doctor in this experimental setup considering the readily available computation power but it is capable of complete diagnosis on a high-end computing resource with minimal setting change.

Proposed disease detection system was refined in terms of QoS parameters by using ICN, as this integration enhance quality of service for all the identified parameters without the need of modification in the system modules.
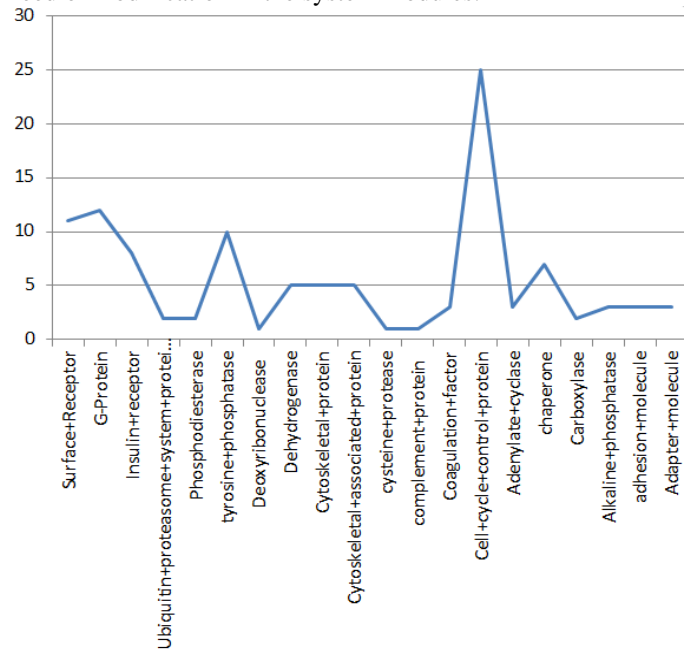

Figure 3. Number of genes present in a protein class responsible for causing cancer.
Figure 4.

V. CONCLUSION AND FUTURE SCOPE

The proposed framework provides promising alternative of conventional disease detection as a gene-based disease detection recommender system which can evolve with use and increase accuracy by incorporating advanced Machine

learning, ANN and Deep learning techniques. Protein class prediction becomes cost effective and usable in disease diagnosis by doctors for more accurate detections leading to better diagnosis and development of personalised medication.

Limitations

The system was implemented for gene-based disease detection recommendations. But some diseases are not genetics based but life style oriented so this system will take time to compensate for that drawback but eventually it will correct itself with its self-learning architecture. Literature review suggests that even life style induced diseases have a risk factor connection with genomic background of an individual. Thus, this limitation is not a deterrent to the proposed framework.

Future work

This framework is designed to be implemented on big data systems but here implemented on a small scale, so it can be applied to big data system. Also, the drug discovery part of framework was not implemented so that can be incorporated. Other alternative tests for protein class identification or just amino acid chain extraction can be used for improved accuracy in the initial stages leading to better detection results and more cost-effective solutions.

REFERENCES

[1]   H.She et al., A network basedsystem architecturefor remote medical applications, Asia pacific Network, 2007

[2]   Vergados, D. J., Vergados, D. D., & Maglogiannis, I. (2006). NGL03-6: Applying Wireless DiffServ for QoS Provisioning in Mobile Emergency Telemedicine. IEEE Globecom 2006.

[3]   Gondal, I., Iqbal, M., Woods, M., & Sehgal, S. (2007). Integrated Sensing and Diagnosis -- The next step in Real Time Patient Health Care. 6th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2007).

[4]   Heinzelman, W., Murphy, A., Carvalho, H., & Perillo, M. (2004). Middleware to support sensor network applications. IEEE Network, 18(1), 6-14.

[5]   Henrion, Stéphane & Mailhes, Corinne & Castanie, Francis. (2004). Transmitting critical biomedical signals over unreliable connexionless channels with good QoS using advanced signal processing 2004.

[6]   Hamed, R. Roose, M. Branicki, A. Rubin MD, "T-Recs: Time-aware Twitter-based Drug Recommender System," in 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2012.

[7]   Rung-Ching Chen, Yun-Hou Huang, Cho-Tsan Bau, Shyi-Ming Chen, "A Recommendation system based on domain ontology and SWRL for anti-diabetic drugs selection," Expert System with Applications, vol. 39, no. 2012, pp. 3995-4006, 2012.

[8]   Rui Chen, Michael Snyder, "Systems biology: personalized medicine for the future?," SciVerse Science Direct,Current Opinion in Pharmacalogy, vol. 12, pp. 623-628, 2012.

[9]   Lawal Mohammed Ma'aruf, Mariya Garba "Design And Implementation Of An Expert Diet Prescription System," International Journal of Artificial intelligence and Expert System (IJAE), vol. 3, no. 4, pp. 126-134, 2012.

[10] C. Doulaverakis, GeorgeNikolaidis, A. Kleontas and I. Kompatsiaris, "GalenOWL: Ontology-based drug recommendations discovery," Journal of Biomedical Semantics, vol. 3, no. 14, pp. 1-9, 2012

[11] Asmaa S. Hussein,Wail M. Omar,Xue Li Modafar Ati "Accurate and Reliable Recommender System for Chronic Disease Diagnosis." GLOBAL HEALTH 2012 : The First International Conference on Global Health Challenges. Ed. GLOBAL HEALTH 2012. IARIA, 2012. 113-118.

[12] Kazuyuki Shimadaa, Hidekatsu Takada, Satoshi Mitsuyama, Hideyuki Ban, Hitoshi Matsuo, Hitoshi Otake, Hiroyuki Kunishima, Keiji Kanemitsu, Mitsuo Kaku, "Drug-Recommendation System for Patients with Infectious Diseases." AMIA 2005 Symposium Proceedings. Ed. AMIA. 2005. 1112.

[13] Yin Zhangy; Long Wangy; Long Huy;Xiaofei Wang,"COMER: ClOud-based MEdicine Recommendation." QSHINE ICST August 18-20 (2014): 1-7.

[14] Burasakorn Yoosooka, Suvi Chomchaiyal, "Medication Recommender System", International Conference on Science and Technology 2015, RMUTT, IEEE, 2015.

[15] Sarath Chandra Janga, Mohan Mallikarjuna Rao Edupuganti, "Systems and Network-Based Approaches for Personalized Medicine" Current Synthetic and Systems Biology 2.3 (2014): 1-5.

[16] Amanpreet Kaur, Sawtantar Singh Khurmi, "A Study Of Cloud Computing Based On Virtualization And Security Threats", International Journal of Computer Sciences and Engineering, Vol.5, Issue.9, pp.108-112, 2017

[17] P. Jyotsna, P. Govindarajulu P, "Stage Prediction of Lung Tumor Identification: A Decision Tree Model for Particle Swarm Optimization Algorithm", International Journal of Computer Sciences and Engineering, Vol.6, Issue.7, pp.1363-1372, 2018.

Authors Profile

Mr. Amritpal Singh did Bachelor of Computer Applications from Guru Nanak Dev University, Amritsar in 2007 and Master of Computer Applications (MCA-FYIC) from Guru Nanak Dev University, Amritsar in year 2009. He is currently pursuing Ph.D. and working as an Assistant Professor in Department of Computer Science, Guru Nanak Dev University, Amritsar since 2012. He has published more than 25 research papers in reputed international journals including Thomson Reuters (ESCI & Scopus) and conferences sponserd by UGC, Springer and ACM. His main research work focuses on Genetic Algorithms, Machine Learning, Cloud Computing, Big Data Analytics, Data Mining, Image processing and Natural language processing. He has 8 years of teaching experience and 6 years of Research Experience.

Mr. Sunny Sharma did Master of Computer Applications from Guru Nanak Dev University, Amritsar in year 2008. He is currently pursuing Ph.D. and has worked as an Assistant Professor in Department of Computer Science, Guru Nanak Dev University, Amritsar from 2012-2017. Now working in Hindu College as an Assistant Professor since 2017. He has published more than 35 research papers in reputed international journals including Thomson Reuters (ESCI & Scopus) and conferences sponserd by UGC, Springer and ACM. His main research work focuses on Genetic Algorithms, Machine Learning, Data Mining, Theory of Computation. He has 9 years of teaching experience and 6 years of Research Experience.

Dr. Gurvinder Singh did Ph.D. and working as a Professor & Dean Faculty in Department of Computer Science, Guru Nanak Dev University, Amritsar. He has published more than 80 research papers in reputed international journals including Thomson Reuters (SCI, ESCI & Scopus) and conferences sponserd by UGC, Springer and ACM. His main research work focuses on Parallel computing, Genetic Algorithms, Machine Learning, Cloud Computing, Big Data Analytics. He has 23 years of teaching experience and 16 years of Research Experience

Dr. Rajinder Singh did Ph.D. and worked as a Professor & Head in Department of Computer Science, Guru Nanak Dev University, Amritsar. He has published more than 55 research papers in reputed international journals including Thomson Reuters (ESCI & Scopus) and conferences sponserd by UGC, Springer and ACM. His main research work focuses on Genetic Algorithms, Machine Learning, Cloud Computing, Big Data Analytics, Data Mining,. He has 20 years of teaching and 16 years of Research Experience

Ms. Amanpreet Kaur did Bachelor of Technology(CSE) from Punjab Technical University in 2012 and Master of Technology in Computer Science and Engineering from Guru Granth Sahib world University, Fatehgarh Sahib in year 2014. She is currently pursuing Ph.D. in Department of Computer Science, Desh Bhagat University, Mandi Gobindgarh.. She has published more than 6 research papers in reputed international journals . Her main research work focuses on Cloud Computing, Big Data Analytics and Machine Learning. She has 1 years of teaching experience and 3 years of Research Experience