Karnataka Jnana Aayoga
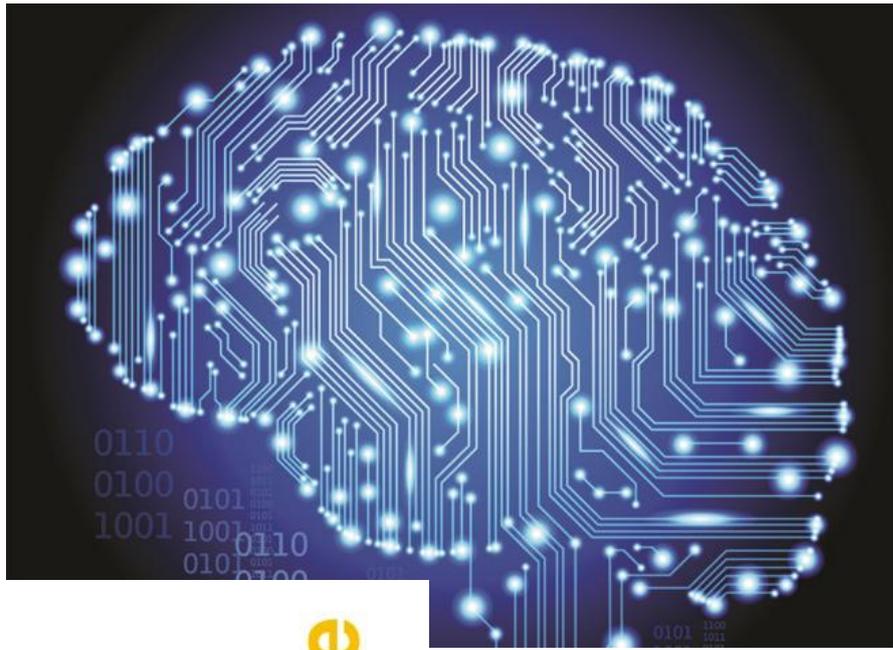## (Karnataka Knowledge Commission)

Report of the Workshop on Machine Intelligence & Data Science Analytics in Education in Karnataka: A Dialogue among Stakeholders (Industry, Academia & Government)

In collaboration with Department of IT, BT and Science & Technology, GoK

Hosted by Indian Institute of Science, Bengaluru

(Report by KJA Task Group on Machine Intelligence)

[Jan 18, 2016]

Table of Content

Abstract

This is a report on a workshop conducted by KJA Task Group on Machine Intelligence. This workshop was held on 18th January 2016 in Computer Science and Automation (CSA) department in Indian Institute of Science (IISc). The aim of the workshop was to explore how Machine Learning and Data Science can be helpful in the social sector, and to deliberate on the undergraduate curriculum for Computer Science in engineering colleges of Karnataka, to add adequate focus on Machine Learning and Data Science. The attendees consisted of faculty and students from CSA department of IISc, faculty members from several engineering colleges in Karnataka, and eminent scientists from industrial research labs in Bangalore.

Acknowledgements

About Karnataka Jnana Aayoga (KJA)

Karnataka is emerging as a Knowledge State in the country and needs to prepare itself to be a knowledge society – through education, innovation, skill development, social development and economic growth. With a view to address these and develop a progressive society in Karnataka state, Government of Karnataka constituted Karnataka Jnana Aayoga. The present KJA has been formally reconstituted in December 2013 under the Chairmanship of Dr. K. Kasturirangan. KJA has 32 experts and department Secretaries as Members. The main aims and objectives of the present KJA are to recommend actions for institution building, policy innovation and excellence in the field of education, health, S&T, industry, entrepreneurship, research and innovation, traditional knowledge, agriculture, E-Governance, rural development and ANY other relevant areas.

Tasks of KJA are mainly "proof-of-concept" and get defined/formulated, either through internal discussions within KJA – mainly issues of public/societal/technological and knowledge relevance for the state OR are identified through interactions with GOK departments – mainly issues of governance and development in the state. KJA tasks are "anchored" with one or more departments of GOK – so that after proof-of-concept stage by KJA, any executive implementation can get effectively coordinated by relevant departments of GOK. As part of "proof-of-concept", KJA takes up technical assessments, studies, small demonstrations, proto-typing analysis, policy analysis, social impact assessment etc – which result in relevant recommendations to GOK.

Karnataka is one of the leading states in technology and machining with a wide range of government, private sector and academia involved in a range of technology – including, machines, information technology, aero-space technology, education technology and a host of industrial technologies. Given this status, Karnataka aspires to rise to greater heights so that its citizens can be assured of advanced technology development benefits and the state can position a futuristic technology-climate that will prepare for an advanced, mature and economically prosperous state and national development processes.

Recent developments in smart mobile systems and cloud computing are enabling the connection of millions of devices and that has amplification of any governance, technological, efficient, moral and ethical imperative through a network of machines. Smart Machines are now becoming a cognitive, contextually aware computing system capable of making decisions without human intervention – that is the way of future. Machine Intelligence technologies – 3d Printing, indigenous Cloud design, Internet-Of-things (IOT), robotics, Artificial Intelligence (AI), algorithmic societal tools, social analytics etc are getting integrated into education systems, high-end industrial capability, analytical governance and citizen-centric services of improved quality of life through efficient governance.

To develop a Machine Intelligence Plan for Karnataka, KJA decided to constitute the **Task Group on Machine Intelligence** composed of technology specialists/experts. The Task Group was formally constituted in November 2015 under the Co-Chairmanship of Mr. Venkatesh Valluri and Prof. Chiranjib Bhattacharya.

The main task of the KJA-Task Group on Machine Intelligence is to prepare a strategy for Karnataka to take up leadership in development of Machine Intelligence technologies through an integrated development of key technologies, applications development, education and research and industrialistion. The TG will also define a roadmap for development and industrialisation of

technologies and applications of Machine Intelligence for the benefit of the society, state and nation-specifically in areas of healthcare, science & technology education, transportation, rural & urban development, governance, citizen services etc.,

About the Workshop

To achieve its objective, the TG has planned for a series of stakeholders consultations, in partnership with Department of IT, BT and Science & Technology, and one such activity is the 'Workshop on Machine Intelligence and Data Science Analytics in Education in Karnataka: A dialogue between Stakeholders' held on January 18, 2016 with support from Department of CSA, Indian Institute of Science, Bangalore.

The aim of the workshop is to engage with all stakeholders including Government, Academia and Industry, and discuss the state of Machine Learning and Data Science Education in the state of Karnataka. More specifically, the aim is to understand the quality of current engineering graduates and to identify specific government interventions which would help improve the quality of Education. The workshop specifically aimed at identifying existing gaps in our academic system by bringing together three key stakeholders: Industry-Academia and Government for a dialogue. The workshop captured the perspectives of all the stakeholders on 'What is Machine Intelligence'; Where is the Gap? and What needs to be done to make Karnataka a lead in the Machine Intelligence technology.

The organizers of the workshop were Prof **Chiranjib Bhattacharyya**- a professor in CSA department who specializes in Machine Learning, Dr. **Mukund Rao** – Member Secretary, Karnataka Jnana Aayoga, Mr. Venkatesh Valluri, KJA Member and Co-Chair of the TG, and Dr. **Maulishree Agrahari** – CEO, ICT Skills Development Society, Government of Karnataka. It was attended by several luminaries from Bangalore-based industries who are working in the areas of Data Science. They included Dr. **Manish Gupta** – Director, Xerox Research Center India, Dr. **Ramesh Hariharan** – CTO, Strands Life Sciences and Dr. **Ravi Kannan** – a data scientist of worldwide repute who is currently associated with Mircosoft Research India and is also an adjunct faculty member in CSA department of IISc. It was attended by several faculty members from various engineering colleges in state. Several members of the Karnataka Jnana Ayoga and Department of IT, BT, S &T under Government of Karnataka were present, as were several students of CSA department. The total audience size was about 50 (Annexure-II: Participants List).

The workshop started by a welcome address by Prof. Jayant Haritsa, chairperson of the Department. He welcomed the delegates, and said that the department will be happy to support Karnataka government's efforts to strengthen education and research in Data Science and Machine Learning, and to put these use in governance.

What is Data Science?

According to Wikipedia, "Data Science is an interdisciplinary field about processes and systems to extract knowledge or insights from data in various forms, either structured or unstructured". Data Science has become a buzzword in the Information Technology industry over the last few years. Three factors have greatly contributed to the growth of this field: 1) Improvement in sensing technology that allows us to collect data from various sources- from consumer behaviours on online shopping portals to blood pressure readings of patients in hospitals 2) Improvement in computing and storage technology that facilitates efficient storage and analysis of Terabytes of data 3) Improvement in Data Mining/Machine Learning research resulting in efficient algorithms for data analysis, discovering patterns or trends, and building predictive models for future measurements.

Data science research is very much interdisciplinary in nature. We need researchers from the data domain (eg. Climatologists for weather/climate data) who can understand the data and ask the right questions, computer scientists and data mining experts who can do the analysis using existing or novel methodology and finally the policy-makers, who can interpret the analysis results and use them to frame new policies. The first two stakeholders can be from both academia and industry.

Session-2

Industry Perspective on MI and Data Science Curriculum-The Necessary and the Sufficient

Several private-sector companies in India have started to focus heavily on Data Sciences in recent years. So, representatives from three such companies were invited to present their perspecives on these matters, and give the audience snapshots of how they utilize Data Science in their work.

The first person to speak was **Dr. Manish Gupta** from **Xerox Research Center, India**. His talk was roughly on two topics- education and healthcare in India, and how Data Sciences can play a role in them.

Regarding education, Dr. Gupta prominently talked about how the quality of regional engineering colleges in India are poor in terms of teaching and employability of their alumni. He felt that premier institutes like IISc and the IITs, which have some of the best engineering teachers in the country, should "democratize" their course materials through online courses using frameworks like MOOC (Massive Open Online Courses) which can allow students of less privileged colleges to access them. And Xerox Research Center India is trying to contribute here by building technologies for easy management of such online course contents, that will allow teachers to efficiently compile teaching material from multiple online resources, and the students to browse these online contents as efficiently as reading a well-indexed book. He also discussed repeatedly the need of personalization of such content, so that each student can use them according to his/her personal requirements. For example, if some students are more comfortable in studying in a local language, machine translation and natural language processing techniques can help them generate it.

The related technologies developed by his lab that he specifically mentioned are:

1. Multisource content composition for planning an online course by instructors.

2. Automatic table of content generation from an online course video, for easier browsing by students.

3. Generation of word cloud from video, which will provide students with a summary without watching it fully

4. Evaluation of teaching style- how engaging the video lectures are with the viewers!

In the second part of his talks, Dr. Gupta focussed on healthcare problems. He pointed out how the monitoring of daily activities by smart devices can help in preventing diseases. Analysis of clinical data can be used to build predictive models for various health indicators of a patient, that will allow the hospital authorities to shift them into and out of ICU at appropriate times. This is important since ICUs are very costly.

Dr. Ramesh Hariharan is the founder and chief technical officer of **Strands Life Sciences**, a company that develops computational and analytical techniques for life sciences. The company specializes in data mining and predictive modelling of genomic data to aid clinical research. He has also been teaching a course called Data Analytics in IISc, that serves as an introductory course to Data Sciences for undergraduate and graduate students.
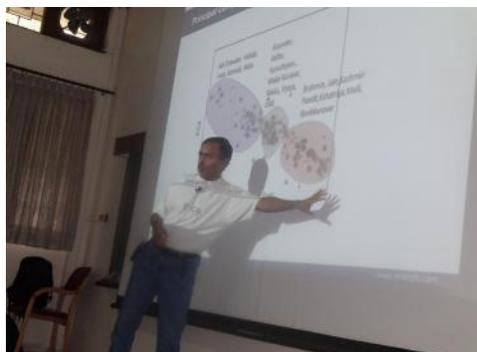

During his presentation, Dr. Hariharan enthralled the audience with an analysis of the Indian caste system through genomic data collected from a large number of people across various castes. By using a data analysis technique called Principal Component Analysis, he showed how the data gets partitioned into clear clusters, and it turns out that the data-points within each cluster are primarily from the same caste. This indicates that people have traditionally married within same caste which is reflected by caste-specific genomic patterns. On further analysis of each cluster, smaller sub-clusters emerge, and it turns out that each corresponds to a sub-caste! Thus, phenomena which are well-known by sociological analysis can be verified by appropriate analysis of genomic data.

Dr. Ravi Kannan from **Microsoft Research India** is a leading Computer Scientist of world renown, and has worked on many theoretical problems that form the basis of Data Science. Most of his work has been theoretical, involving advanced mathematics.

In the workshop, he spoke about the relations between Computer Science, Data Science, Machine Learning, and the mathematical concepts that are at their core. In very simple language and using simple examples he illustrated some concepts in Data Science such as clustering, and under what situations standard clustering techniques such as K-means work and when they fail. He went on to talk about the undergraduate curriculum, pointing out that it is very important for students of Data Science to know Probability, Linear Algebra and High Dimensional Geometry well. He pointed out that Computer Science curriculum generally harps on Discrete Mathematics, as that forms the basis of the more traditional theoretical computer science. However, Data Science is built more around continuous mathematics, and the curriculum should be revised to change the focus. However, he also said that the mathematics should be taught in conjunction with Computer Science, not as a disconnected subject.
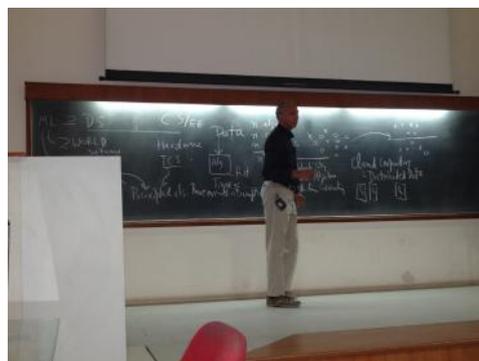
Session-3

Students Perspective on MI and Data Science Curriculum: "What's Correct and what's Missing"

In the post-lunch session, some former students of the department, who had pursued doctoral research in Machine Learning and have recently submitted their theses, were invited to present their opinions about possible curriculum changes for undergraduate degree program in Computer Science. These students were Adway Mitra (who had done B.E. in Computer Science from Jadavpur University, Kolkata and currently a Research Associate in Divecha Center for Climate Change, IISc), Arun Rajkumar (B.E. from P.S.G. College of Technology, Coimbatore, currently with Xerox Research Center) and Prashanth L. A. (B.E. from NITK Surathkal, currently postdoctoral fellow in University of Maryland, USA).

All three students stressed on the importance of teaching Linear Algebra and Probability at the undergraduate level, since these are the building blocks of Data Science. Both Adway and Prashanth pointed out that these subjects are currently part of the curriculum, but are often neglected by students as they fail to see any connection with Computer Science. A proposal was to illustrate various concepts in these subjects with their applications in Data Science.

Additionally, an elementary course to introduce the basic concepts of Data Science and their uses in various fields, as well as a lab course for hands-on experiments with data from various domains were proposed. Arun proposed that undergraduate students should be provided exposure to industries and research labs through internships and summer/winter schools.

Session-4

Engineering College Representatives Perspectives on Current Contents of Machine Learning, Machine Intelligence & Data Science-UG & PG Curriculum

Representatives from several engineering colleges in Karnataka presented the state of affairs in their colleges regarding training in Data Science. They presented statistics of faculty members and students, details of course program related to Data Science and details of research related to Data Science. The colleges which were represented were: R.V. College of Engineering (Bangalore), B.M.S. College of Engineering (Bangalore), B.M.S. Institute of Technology and Management (Bangalore), P.E.S. Institute of Technology (Bangalore), Gogte Institute of Technology (Belagavi), Malnad College of Engineering (Hassan), S.D.M. College of Engineering and Technology (Dharwad) and Sri Jayachamarajendra College of Engineering (Mysore).

Courses and labs: From these presentations, it was found that most of the colleges have started to offer courses related to Data Mining, Information Retrieval, Big Data Analytics, Natural Language Processing, Image Processing etc – all very relevant to Data Science. Some of the colleges are offering MOOC-based courses also. It also turned out that several colleges either have already started or are planning to start research labs and research programs related to Internet of Things, Image Processing, Sensor Networks, etc.

Projects: Representatives of the high-ranked colleges in Bangalore revealed that several B.E./B.Tech. and M.E./M.Tech. Projects in recent times have been related to Data Science. The most prominent and socially relevant ones included

1) Sentiment analysis from text data from internet (BMS)

2) Emotion detection from images and videos (BMS)

3) Vehicle identification and number-plate recognition in videos recorded at traffic signals (BMS)

4) Analytics on medical data and sensor data collected from wearable devices (BMS)

5) Price prediction for agricultural products (SDM Dharwad)

6) Recommendation systems for transport (SDM Dharwad)

7) IoT-based irrigation systems (SDM Dharwad)

8) Landcover classification (PESIT)

9) Detection of accidents, fights etc in CCTV surveillance videos (RVCE)

10) Modeling and analysis of telecom networks (RVCE)

Shortcomings: However, different colleges seem to have achieved different levels of success with respect to these initiatives. The colleges in the smaller towns like Belagavi, Dharwad etc complained that they do not have enough well-trained faculty members to teach these course, available teaching materials such as books are of very low quality, and there is insufficient interest among the students in taking up these subjects and using them for solving practical social problems.

Participating Institutes

- R.V. College of Engineering, Bangalore
- B.M.S. College of Engineering, Bangalore
- B.M.S. Institute of Technology and Management, Bangalore
- P.E.S. Institute of Technology, Bangalore
- Gogte Institute of Technology, Belagavi
- Malnad College of Engineering, Hassan
- S.D.M. College of Engineering and Technology, Dharwad
- Sri Jayachamarajendra College of Engineering, Mysore

Session-5

Panel Discussion & Interaction: Reinforcing Academia to Leverage MI and Data Science for Societal Good

The last event of the workshop was the panel discussion between the main speakers. The panel consisted of Dr. Manish Gupta, Dr. Ramesh Hariharan, Dr. V. Vinay and Dr. Ravi Kannan from the industry, Dr. P. S. Sastry – a faculty member of IISc and Dr. Maulishree Agrahara representing the government. The session was moderated by Prof. Chiranjib Bhattacharyya.

The session started with Prof Chiranjib asking the panellists whether they thought that the present undergraduate curriculum was fine, or whether it needed slight changes, or whether it should be radically altered. All of them agreed that significant, if not fundamental, changes are required. Several specific points were raised about the nature of changes required.

Regarding students, Dr. Manish Gupta pointed out that their quality and employability needs to improve, and they should be excited about working on real problems of the society, through their expertise of technology in general and Data Science in particular.

Regarding subjects and courses, Dr. Manish strongly endorsed the suggestions made earlier by the three PhD graduates regarding introductory and lab courses, and about integrating Mathematics courses with Computer Science courses. Dr. Ramesh Hariharan mentioned the need of reducing the number of courses, build the fundamentals strongly and encourage students to learn from hands-on experiences rather than from courses. Drs V.Vinay and Ravi Kannan mentioned the need of teaching continuous mathematics along with discrete mathematics, because Data Sciences need more of the former unlike traditional Computer Science.

Regarding quality of teachers and teaching materials, Prof. P.S. Sastry felt that their inferior quality was the problem rather than curriculum, and warned against arbitrary curriculum changes, because if new contents are to be introduced, they must come in at the cost of some existing contents. Several college representatives pointed out that the number of teachers is not a problem, but their quality is. Dr. Manish, Dr. Maulishree and several members of the audience pointed out the need of creating modernized course content using frameworks such as MOOC, the need for leading institutes (which do have good teachers) to make their classroom lectures accessible to other colleges, and conducting enrichment programs for teachers. Dr. Maulishree mentioned Faculty Development Programs being envisioned by the government. Dr. V.Vinay also pointed out that in many cases even available online lecture videos from NPTEL are not particularly

popular among students, because they may often appear boring, as they don't engage the students.

As **<u>closing comments</u>**, all the panellists lauded the government's initiative to engage with academia and promote Data Science, and wished that some actionable plan comes out. They also applauded that several colleges are paying heed to students' aspirations and trying to introduce course material related to Data Science, albeit with varying degrees of success. Dr. Gupta reiterated the need to use modern technology to access quality course materials from the best institutes and to form networks of colleges for this purpose. Dr. Hariharan repeated that the government should open up data to the public domain, and engage researchers from industry and academia to work on problems of social interest. Dr. Maulishree exhorted the college teachers to teach with passion and write their own textbooks if necessary.

Workshop Summary

The salient points which arose from the entire workshop are as follows:

1) *There is ample scope of using Data Science and Machine Intelligence in governance and policy-making.*

2) *Some industries based in Bangalore are using Data Science successfully in the fields of biotechnology, genetics, education and healthcare.*

3) *It is important for students to build a comprehensive background in continuous mathematics/calculus, probability and linear algebra during their undergraduate studies. The existing courses in these subjects often appear unconnected from Computer Science/Data Science, the connections must be made clear to the students when they are taught.*

4) *It is also important to provide undergraduate students of Computer Science with hands-on experience of data analysis through lab courses and/or internships.*

5) *Several engineering colleges are already trying to give their students exposure to Data Sciences through course related to Data Mining, Machine Learning, Natural Language Processing etc. However, the quality of textbooks remains a concern.*

6) *It would be useful if the students get exposure to topics mentioned in Annexure I-Partial List of Topics.*

7) *Several B.Tech and M.Tech projects related to Data Science done recently by students of these colleges have social relevance.*

8) *Some of the colleges in Bangalore are even in the process of setting up laboratories and research centers dedicated to research on Data Mining and Machine Learning applied to certain fields.*

9) *Faculty and students of CSA department in IISc will be happy to collaborate with Karnataka Jnana Ayoga in their initiative regarding Data Science.*

Annexure-I Partial List of Topics

- **Mathematical Background**
  - Convex Optimization
  - Gradient Descent
  - Stochastic Gradient Descent

- **Data Exploration**
  - Principal Component Analysis
  - Decision Tree
  - SVM
  - Random Forest
  - Matrix Factorization

- **Inference**
  - Belief Propagation
  - E-M algorithm
  - Gibbs Sampling
  - Variational Inference

Annexure-II-List of Participants

1. Mr. Venkatesh Valluri, KJA Member and Co-Chair, TG-MI
2. Prof. Chiranjib Bhattacharya, Co-Chair, TG-MI
3. Prof. Ramesh Hariharan, Member, TG-MI
4. Prof. Ravi Kannan, Member, TG-MI
5. Dr. A. Maulishree, Member Secretary, TG-MI
6. Dr. Shantha Rangaswamy, Associate Professor, Dept. of CSE, IISc
7. Dr. Manish Gupta, Xerox
8. Dr. M.N. Vijayalakshmi, Associate Professor, Dept. of CSE, IISc
9. Dr. Ravinder Prakash, Dean, BMSIT
10. Dr. M.C. Hanumantharaju, Associate Professor, BMSIT
11. Dr. Kishore Reddy Konda, Post Doc, University of Frankfurt
12. Prof. S.R. Biradar, Professor
13. Prof. S. S. Saraf, Associate Professor
14. Prof. K.V. Subramaniam, Professor, PES University
15. Prof. V. Ravishankar, Assistant Professor, MCE, Hassan
16. Prof. K .S .Praveen, Assistant Professor, MCE, Hassan
17. Prof. P.S. Sastry, IISc,
18. Dr. P. Ramakanth Kumar, RVCE
19. Prof. Partha , CDS & CSA, IISc
20. Dr. B. Ramesh, Professor, CSE & MCE, IISc
21. Prof. N. Viswanadham, Professor, INSA, CSA, IISc
22. Prof. Shanta Rangaswamy, Associate Professor, RVCE
23. Dr. Gowrishankar, BMSCE
24. Dr. L.A .Prashanth, CSA, IISc
25. Prof. Chandra R Murthy, ECE, IISc
26. Dr. H.S. Guruprasad, Professor & Head CSE, BMSCE
27. Dr. Kamala, PEST
28. Ms. N. Vidya, Ph.D Scholar, PES University
29. Mr. Harit Vishwakarm, Student, CSA, IISc
30. Ms. Lavanya, Ph.D, CSA, IISc
31. Mr. Lucky Dhakad, MECSA
32. Mr. Sayantan Mukherji, Ph.D, CSA, IISc
33. Mr. Arun Rajkumar, Ph.D, IISc
34. Mr. S. Raman, Ph.D, IISc
35. Mr. Adway, Ph.D, IISc
36. Mr. Piyush, CSA, IISc
37. Mr. Nirmalendu, CSA, IISc
38. Mr. Soham Pal, CSA, IISc
39. Ms. Aashreye, CSA, IISc
40. Mr. Tarun Choubisa, ECE, IISc
41. Mr. U. Raviteja, ECE, IISc
42. Mr. Ankit Jauhari, CSA, IISc
43. Ms. Sarmitha Dutta, CSA, IISc

44. Mr. P.V. Kumar, CSA, IISc
45. Mr. Ashish Srivastava, Student at CSA, IISc
46. Ms. Jaya Gopalkrishna, Research Scholar, PESIT
47. Mr. K.R. Ramakrishna, EE, IISc
48. Ms. Geetha, Dept. of ITBT
49. Ms. S.K. Padma, CSE, IISc
50. Mr. Jay Oza, CSA
51. Mr. Deepak, SRA, KJA
52. Dr. B.S. Padmavathi, Convenor, TG-MI