



PhUSE US Connect 2019

Paper TT02

Voice Based Clinical Monitoring Assistant

Anoop Ambika, Genpro Life Sciences, Thiruvananthapuram, India
Ajith Baby Sadasivan, Genpro Life Sciences, Thiruvananthapuram, India

ABSTRACT

Clinical trials often produce large amounts of data. This ongoing data can be very useful for deriving intelligent insights about the trials. Sponsors can monitor the Safety, Efficacy, Data Quality and Risk Indicators of a clinical trial by effectively visualizing and analyzing data.

That being the case, clinicians are often faced with the challenge to quickly and effectively analyze the data and identify patterns across multiple sites, cohorts and demographics. This has led to an increased adoption of data visualization tools in clinical trial data analytics. In this process, data is extracted from the EDC systems to create clinical data warehouses, which are then used by the clinicians to derive intelligent insights on current and future trials. Some of the commonly used tools include TIBCO Spotfire®, Tableau and QlikView.

Monitors always rely on a visualization team for data transformations and dashboard creation before they could get sufficient answers for their questions. These dependencies result in significant delays and costs, while analyzing the data.

This paper presents MOSS, which is a dynamic software framework powered by natural language processing and machine learning algorithms. MOSS converts a natural language question, from the clinician, (e.g. What are the treatment emergent adverse events in site 1) to a structured query and further visualization. This eliminates the dependency on data teams and allows clinicians to analyze on their own with a simple voice/text command interface. It also enables further drill down features so that users can explore data, based on the output that gets generated from the question. Our state-of-the-art voice interface will help the clinician to simply ask the question and get the answer. As the natural language model matures, we expect MOSS to be the one stop voice assistant for all clinical trials across therapeutic areas.

INTRODUCTION

MOSS aka "**M**onitoring **A**ssistant" is a voice-based application to visualize and analyze clinical trial data with a simple command. MOSS is a web-based application which will combine the features of CDISC, Natural Language Processing and Machine learning. The scalable SAAS platform allows multiple users to collaborate on the same trial to analyze and visualize.

MOSS offers connector modules by which you will be able to import the data from EDC systems. Once the data source is selected and processed, user will be able to ask questions in plain English and the AI layer of MOSS, powered by ML/NLP algorithms will figure out the question intent and provide you with the answer. The output will be further rendered in an aesthetic format to the users. It also allows users to create and save dashboards and provide data drill down capabilities. MOSS supports a variety of Table and Graph presentation formats for the user to choose from. It provides a voice interface for user to speak into it.

MOSS offers a trial management module to create trials and connect the EDC databases to the trial space. Statistical capabilities are also built-in which will help the clinicians perform ad-hoc analysis. Clinicians can create dashboards for different sites, trials, patients etc. MOSS maintains a question history so that users get recommendations on what to see next.

Potential users include Clinical Data Monitors, Clinical Data Management, Biometrics, Regulatory Affairs and Remote Monitoring teams. MOSS eliminates the requirement to learn a data visualization tool and augments the users with a simple natural language interface.



OBJECTIVES

MOSS will be a simple natural language based clinical data visualization interface which aims at simplifying the clinical data visualization and analysis process for the clinical data professionals. Some of the applications which MOSS can find in clinical data management include the following.

DATA VALIDATION

Data validation is the process of testing the validity of data in accordance with the protocol specifications. Edit check programs are written to identify the discrepancies in the entered data, which are embedded in the database, to ensure data validity.

MOSS provides libraries to compute descriptive statistics over your clinical data. These statistics are used to help data managers investigate and understand their data, as well as infer a schema. MOSS have defined connectors to multiple EDC vendors, and to provide flexibility and scale. MOSS can be used to analyze and validate data at scale. One of the use cases of MOSS is the validation of continuously arriving data and skew detection. The inferred schema can be used to parse data into a pre-processing function. MOSS can detect distribution drifts between successive versions of the data dumps. If drift is detected, an appropriate message is supplied to the user. Data skews referring to a difference in the distributions between the data aggregated in specific intervals can also be detected. In addition, MOSS can visualize the statistics of data dumps at specific intervals and thus highlight potential errors or drifts.

SAFETY ANALYSIS

Drug-development processes must ensure patient safety as a sole priority during and after clinical trials. Monitoring of patient safety at all levels of drug-development is hence given utmost importance.

MOSS offer safety analysis question sets for Data Monitoring Committees and Remote Data Monitors. Users will be able to analyze the safety data by various parameters like sites, patient cohorts, severity and a lot more parameters. Users will be able to ask questions which could either be classified into one of the existing questions or new questions can be added in a matter of minutes using an interface called Quantico that MOSS provides.

Patient narratives are also provided as part of the safety analysis. Since patient narratives can be extremely complex to program with data coming from multiple data sets, MOSS becomes an easy to use tool to generate patient narratives in different formats.

Some of the safety analysis questions include

- Display all Treatment Emergent Adverse Events
- Display the most frequent Adverse Event
- Show frequency of mild, moderate and severe adverse events
- All reported treatment emergent headache with VITALS
- Display the narratives for XXXX
- Show the efficacy narratives for all subjects who had a serious adverse event

PHARMACOVIGILANCE

Pharmacovigilance teams are often in charge of ensuring patient safety, adhering to regulatory standards. A state-of-the-art technology infrastructure is important to achieving this goal. MOSS automates the patient data processing and other pharmacovigilance activities. MOSS employs Artificial Intelligence and Natural Language Processing algorithms in order to build recommendation engines that will provide the necessary analysis at the fingertips of the team.

SIGNAL DETECTION

Signal detection helps the pharmacovigilance and monitoring teams to identify any new risks associated with a drug or explode any changed in the risk indicators during the course of the trial. Signal detection can happen through the following activities

- Realtime reporting
- Data monitoring systems
- Clinical trials
- Epidemiology Studies.
- Animal Trials
- Literature Review
- Meta-Analysis
- Other sources



MOSS can connect to multiple EDC systems and to drug reaction databases such as Eudra-vigilance and the FDA Adverse Event Repository system. A large number of adverse events associated with a drug's use compared with the anticipated rate is called a safety signal. Signals can arise at any time during the life of a drug starting from pre-clinical phase to post-marketing phase.

MOSS can generate signals through the comparison of the number of AE's observed in a population against an expected number. If a safety signal is identified, the signal is further processed to see whether there is any relation between the signal and the drug. Cases are further examined for biologic plausibility and potential confounding factors. MOSS can help you find the excess occurrences of adverse events across the trial, sites or cohorts. It can also find out the relationship between the drug and a signal and also provides a statistical interface to perform hypothesis testing to examine the biological plausibility.

CLINICAL ANALYSIS

Clinical Analysis is paramount while performing the efficacy analysis. Clinical questions can be categorized as either background or foreground.

Background questions ask for general knowledge about a condition, test or treatment. These types of questions typically ask who, what, where, when, how & why about things like a disorder, test, or treatment, or other aspect of healthcare.

Foreground questions ask for specific knowledge to inform clinical decisions. These questions typically concern a specific patient or population. They tend to be more specific and complex than background questions.

Quite often, foreground questions investigate comparisons, such as two drugs, or two treatments. MOSS provides you the capability to handle both types of questions.

e.g. Display all WHITE subjects with a history of OSTEOARTHRITIS along their BMI and AGE.

REMOTE MONITORING

With the increased complexity of clinical trials, there comes an additional responsibility to effectively monitor these trials.

MOSS helps the remote monitoring teams identify trends and potential problems that go beyond source documents. MOSS will help the Centralized Remote Monitors to spot the discrepancies that might happen across the sites. It helps problems to be identified and resolved in real time, significantly reducing costs.

To use remote monitoring effectively, a risk assessment must be performed and included in the monitoring plan prior to the start of any clinical program and adapted as the program evolves.

MOSS helps the monitors to determine the 'data quality', a process known as Data Quality Oversight. MOSS helps the monitors with visualizations that use statistical algorithms to analyze subject and site data to detect erroneous/fraudulent data, as well as any anomalies or outliers. These data quality oversight checks can be applied to any trials regardless of the monitoring preference.

TECHNOLOGY

MOSS makes use of both Artificial Intelligence and Natural Language Processing algorithms in order to help the user to ask what they want about the clinical data beneath and get a real time concrete visualization and analysis of data. Some of the technology modules of MOSS include the following.

NATURAL LANGUAGE UNDERSTANDING

Natural language understanding (NLU) is a branch of artificial intelligence (AI) that uses computer software to understand input made in the form of sentences in text or speech format. NLU directly enables human-computer interaction (HCI). NLU understanding of natural human languages enables computers to understand commands without the formalized syntax of computer languages and for computers to communicate back to humans in their own languages.

MOSS uses state of the art NLU algorithms to reduce human speech into a structured clinical ontology. AI fishes out things such as intent, databases and parameters to use while querying the data.

NEURAL TRANSLATION

Neural Machine Translation (NMT) is an end-to-end learning approach for automated translation, which is usually used in language translations. MOSS is employing the Neural Translator for translating a natural language clinical query to a structured database query. This model employs an encoder-decoder framework as the underlying translation model, based on Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM). Basically, in



the encoder phase, the neural network identifies and maintains the semantic information of the natural language query. In the decoder phase, it outputs a new sequence in another language based on the information maintained in the hidden states of the neural network.

TRAINING DATA GENERATION

MOSS uses a library called Snorkel to create large amounts of training data in order to get good performance from for the Neural Translator. A key bottleneck in developing MOSS is the need for large, high-quality training sets for each therapeutic area. In Snorkel applications, instead of tediously hand-labeling individual data items, a user implicitly defines large training sets by writing programs, called labeling functions, that assign labels to subsets of data points, albeit noisily.

This idea of using multiple, imperfect sources of labels builds on work in distant supervision. However, if ignored, the uneven (and unknown) accuracies and coverages of the user provided labelling functions can easily lead to suboptimal results

Snorkel addresses this challenge of uneven training source quality by automatically learning a statistical model of the labeling functions' accuracies and correlation structure. The lack of hand-labeled data when learning this model raises several statistical challenges including estimating accuracies, learning correlations, and selecting features that refine labelling function quality. Snorkel then uses this model to combine and reweigh the labeling functions' labels, producing a set of probabilistic training labels, thus effectively passing along key provenance information about the training.

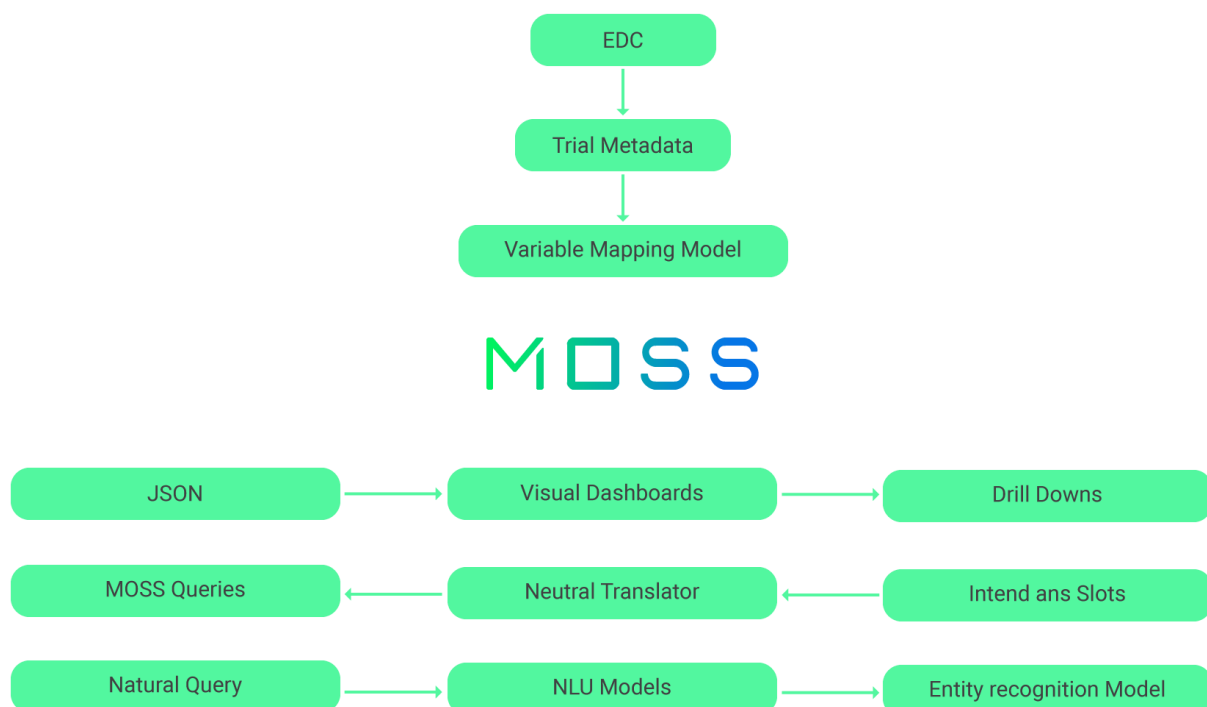
DASHBOARD & DRILL DOWNS

MOSS provides dashboards where users can aggregate the information (charts, tables, tokens) that they collate. Users will have the provision to add or delete answers to a dashboard and every time the dashboard is loaded the latest data will be reflected.

MOSS also provides drill down capabilities which will allow users to click on the outputs and further explore the data. For e.g. all patient related information like, Demographics, Medical History, et al can be obtained by clicking on the patient ID.

MOSS DATA FORMAT

MOSS have developed an algorithm which will import a database schema from multiple EDC vendors and use the labels to create a mapping to a standardized data format called MOSS Data Format (MDF). User will be given the provision to edit the mapping file and once the EDC data is mapped to the MDF, MOSS will use the MDF structure to generate the queries using the Neural Translator.





MODULES

MOSS contains the following software modules which will help the user to create, trials, users, and connect to the EDC databases in the back end.

USER MANAGEMENT

This module helps the users to register themselves in MOSS and the administrator will then accept or reject the user request. Administrator can also assign one or more trials to a user.

TRIAL MANAGEMENT

User will have the provision to create trials and assign users to specific trials. Basic information about the trial will be captured in this module.

DB CONNECTORS

MOSS provides a database connector module which will help you connect to multiple EDC data sets or upload your SDTM or ADAM datasets for further analysis and visualization.

VISUALISATION

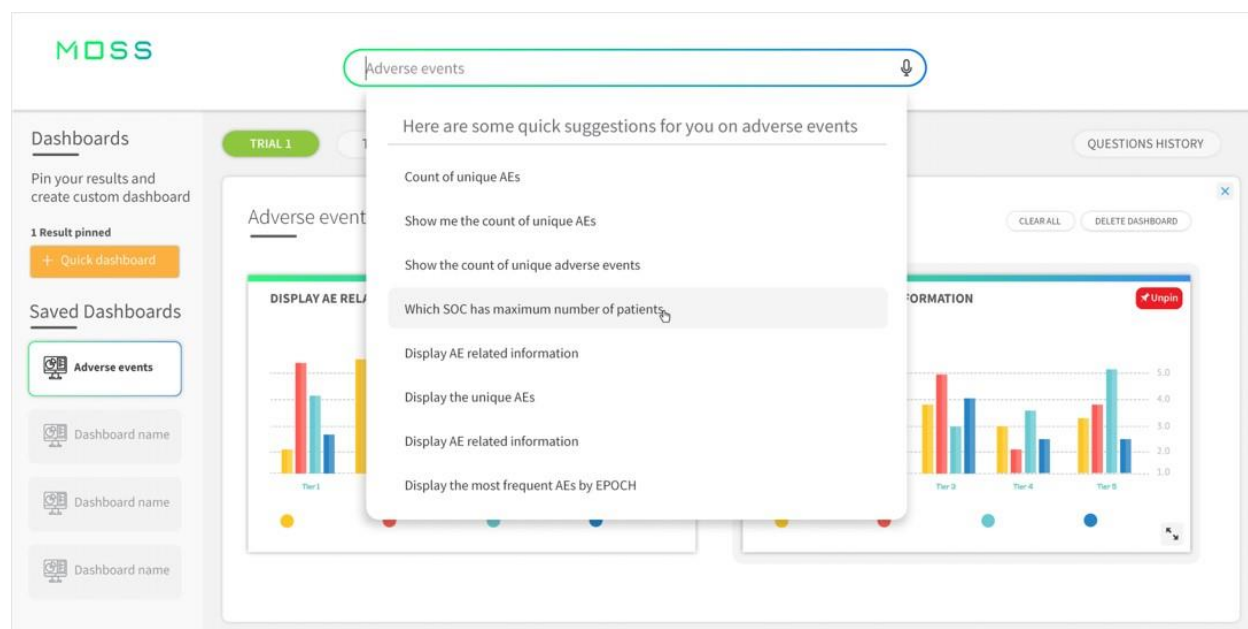
MOSS comes with a powerful visualization module which will allow the users to create the dashboards and also allow further drill down on the data.

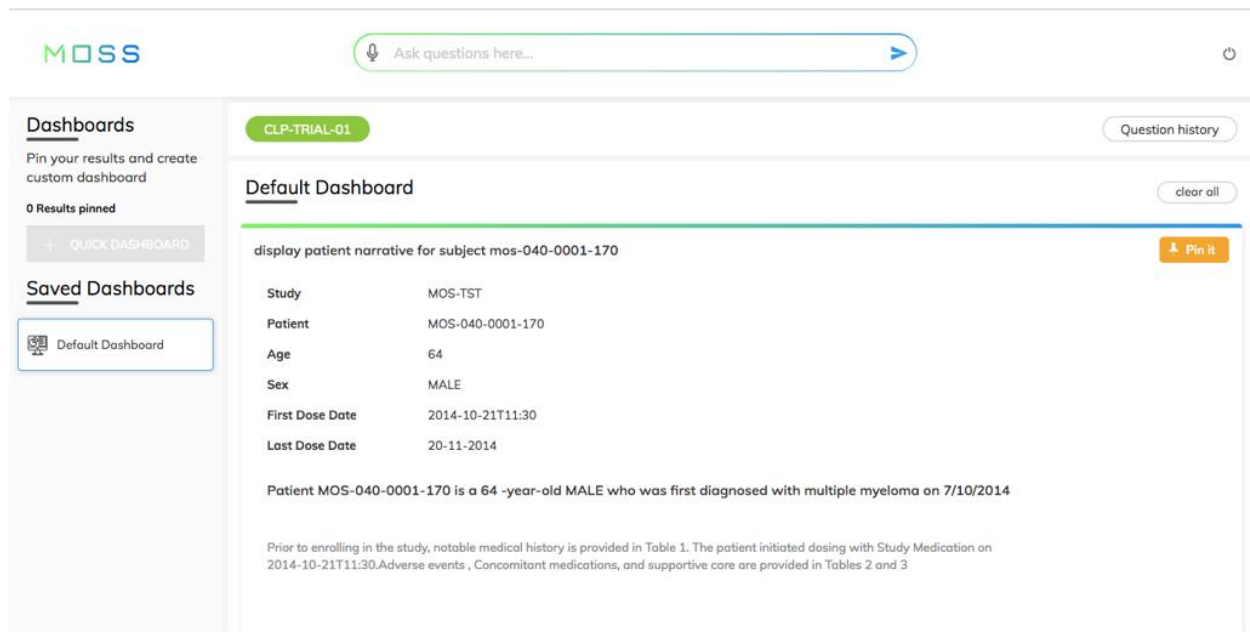
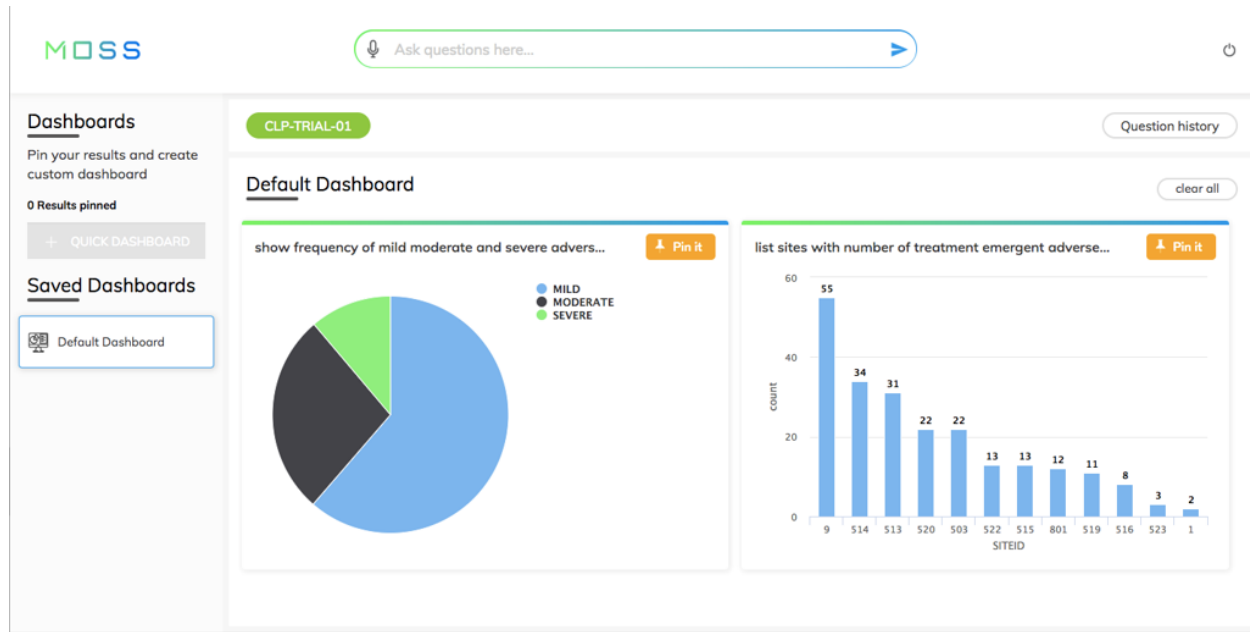
CASE STUDY

To start off with we have considered an open-label single sequence, cross-over drug-drug interaction study, investigating the effect of drug on the pharmacokinetics. The data collected included

- Inclusion/Exclusion
- Demography
- Medical history
- Physical examination
- Height, weight, body mass index (BMI)
- Vital signs (heart rate, blood pressure, tympanic temperature and respiration rate)
- 12-lead electrocardiogram (ECG)
- Urinalysis
- Blood samples for hematology and biochemistry

Some of the sample screens which depicts the analysis of the data is provided below.





FUTURE PLANS

MOSS is currently available as a product ready for early adoption. Some of the areas that we are planning to improve include

- Quick and easy training data generation
- Support analysis on ADAMS
- Statistical Plugins that will allow ad-hoc analysis for the monitor
- Visualization of Data Shift Patterns
- Incorporate better Signal Detection



CONCLUSION

Powerful Voice/Text Interface, Natural Language Queries and Visualization and drill down capabilities make MOSS one of the most useful platform for Clinical Data Visualization. Data Validation, Interim Analysis, Safety Analysis, generation of Patient Narratives etc. Clinicians will find the new interface easy to use and will see their dependency on visualization and data teams slowly getting reduced. This will result in cost savings in terms of resources, license costs and time lines.

REFERENCES

Reading Wikipedia to Answer Open-Domain Questions
Danqi Chen, Adam Fisch, Jason Weston, Antoine Bordes

NLPReViz: an interactive tool for natural language processing on clinical text.
Trivedi G, Pham P, Chapman WW, Hwa R, Wiebe J, Hochheiser H.

Practical Applications for Natural Language Processing in Clinical Research: The 2014 i2b2/UTHealth Shared Tasks
Özlem Uzuner and Amber Stubbs

WIKIQA: A Challenge Dataset for Open-Domain Question Answering
Yi Yang, Wen-tau Yih, Christopher Meek

Snorkel: Rapid Training Data Creation with Weak Supervision
Alexander Ratner, Stephen H. Bach, Henry Ehrenberg, Jason Fries, Sen Wu, Christopher Re.

ACKNOWLEDGMENTS

We would like to acknowledge the contributions of Arjoonn Sharma, Praveen S, Vishakh V S, Vani S and Bhavya K from CLAP Research who worked on the NLU algorithms, Neural Translation Algorithms and user interface for MOSS. We would also like to acknowledge the team of Genpro Life Sciences and Travancore Analytics for helping us with the NLU training datasets.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Author Name : Anoop Ambika
Company : Genpro Lifescience India Private Ltd
Address : Technopark, India
City / Postcode : Thiruvananthapuram, 695581
Work Phone : +91.471.2700151
Email : anoop.ambika@genproindia.com
Web : www.genproindia.com

Author Name : Ajith Baby Sadasivan
Company : Genpro Lifescience India Private Ltd
Address : Technopark, India
City / Postcode : Thiruvananthapuram, 695581
Work Phone : +91.471.2700151
Email : ajith.nair@genproindia.com
Web : www.genproindia.com