ISSN: 2633-4828 Vol. 6 No.1, January, 2024

International Journal of Applied Engineering & Technology

AN AI BASED TECHNIQUE FOR POLL RESULT PREDICTION

M. Suresh Babu[1], A. Pranayanath Reddy[2], D. Asha Devi[3] and B. Deevena Raju[4]

[1]Professor and Dean at Teegala Krishna Reddy Engineering College in Hyderabad

[2]Associate Professor in the Department of CSE- at TKR Engineering College,Meerpet, Hyderabad .

[3]Professor in the Department of Electronics & Communication Engineering, SNIST, Hyderabad

[4]Assistant Professor in Faculty of Science & Technology – ICFAI Tech School,n Hyderabad

[1]sureshcse@tkrec.ac.in, [2]a.pranayanath@tkrec.ac.in, [3]ashadevi@sreenidhi.edu.in, [4]deevenaraju@ifheindia.org

An AI based Technique for Poll Result Prediction, International Journal of Applied Engineering & Technology6(1), pp.1-5.

ABSTRACT

Predicting election results using social media data is indeed a burning topic in politics. With the widespread use of media in political campaigns, it has become increasingly important to understand how to harness the power of social media data for accurate and timely predictions of election outcomes. The proposed MEF framework is an interesting approach to this problem, as it combines both a process and a machine learning model to forecast poll results based on social media performance data. The use of offline polls as labeled data is also a valuable addition to the framework, as it allows for a more accurate and reliable comparison of the predicted results with the actual results. The fact that the framework has been successfully applied to recent Assembly elections in India is promising, as it suggests that this approach may be effective in predicting election outcomes in other similar scenarios. The high degree of precision attained in forecasting the candidates' final vote shares and the capability to make forecasts every day are especially noteworthy since they show the approach's potential to deliver results that are comparable to or even better than those of conventional polls. Overall, the MEF framework represents a promising direction for predicting election outcomes using social media data, and it will be interesting to see how this approach develops and evolves in the future. The approaches and the present issues in this field of study have suggested a novel approach to handling the issues. In similar circumstances, they might potentially be directly used to forecasting next elections.

Index Terms - Machine learning, Precision ,Prediction, Social media

INTRODUCTION

The research topic of predicting election results using social media presents many challenges, as evidenced by several studies and surveys. The most often employed strategy continues to be centered on the volume and sentiment analysis of WhatsApp postings, with the platform being used to survey public opinion. Researchers are advised to employ methods based on various platforms and extra non-SM data, such as polls or economic indicators, as this strategy has been proven to have a poor success rate. The domains include methods, sampling, modeling, performance evaluation, and scientific rigor, have also been provided by the most recent surveys. This work attempts to give a novel strategy for utilizing social media to forecast election outcomes that takes into account these difficulties and provides a more robust and trustworthy method. The method that is still most frequently employed involves the volume and sentiment analysis of social media postings, with the platform serving as a means of gathering public opinion. The typical steps in this approach are as follows: (i) gathering data from WhatsApp using pre-selected keywords over a predetermined time period; (ii) cleaning the data by eliminating duplicates and retweets; (iii) sentiment analysis utilizing lexicon-based or machine learning methods; (iii) prediction employing volume/sentiment counting analysis and a linear formula; (iv) evaluation. The number of posts that mention a candidate and the candidate's vote share are often directly associated, and the performance evaluation might range from merely announcing the winner to closely examining the vote percentage. The most recent surveys have also offered guidance in the domains, including methodologies, sampling, modeling, performance evaluation, and scientific rigor. Since most studies only apply to one election, they highlight the need for: processes that are clearly defined, generalizable, and replicable; a strong technique for segregating data and

analysis due to the arbitrary data collection choices; scientific rigor for the results analysis because previous observations and study state that the election winner or asserting that results were close to the electoral results; and the investigation of the relationship between the processes and the results. Investigate machine learning (ML) techniques and models to handle the particulars of this circumstance rather than merely improving sentiment analysis. Last but not least, anticipating election outcomes may be more accurately referred to as direct casting, which is an assessment of the present or forecast. This is because many people only make up their minds about how they will vote in the last days or even on the day of the election. To address the unique aspects of this situation, it is necessary to develop procedures that are clearly defined, generalizable, and reproducible, as well as a sound methodology for data collecting and analysis. Predicting election outcomes during the campaign is another thing. This study aims to present a novel approach that considers these validations and offers a more robust and reliable method for predicting election results using social media.

A. METHODOLOGY

The current work proposes MEF, a Media system for Political decision, which is a structure intended for anticipating political decision results utilizing web-based entertainment execution highlights and AI draws near. The structure comprises of an interaction and model for Direct projecting political decision results and uses a bunch of measurements in view of Zajonc's composition hypothesis. The ML model utilized for expectation is an outfit of counterfeit brain networks prepared with web-based entertainment measurements as elements and disconnected surveys as marked information. The proposed system was tried on late Get together races in India, beginning from 300 days before the final voting day. The outcomes were genuinely tried with surveyor information from every political decision and contrasted and comparable examinations and verifiable baselines from the surveying business. The outcomes showed that the structure had the option to accomplish an elevated degree of precision in foreseeing the last vote portion of up-and-comers, giving serious or stunningly better outcomes than conventional surveys. The review has endeavored to address a few difficulties in the examination area of electing expectations and has introduced another methodology for assessing citizen inclinations that can supplement customary surveying or be integrated into surveying procedure. The review has likewise researched and applied AI for a new, underexplored setting that presents explicit difficulties. Also, the review has inspected the chance of tracking down nonlinear connections between's internet based conduct and disconnected results. Generally, the review presents a better approach to confront the difficulties of foreseeing political race results utilizing web-based entertainment execution highlights and AI draws near.

DEFINITION OF THE ISSUE AND TECHNIQUES

A Principal Issues Addressed

Training ML algorithms for election predictions can be a challenging task due to the scarcity of historical data and the lack of labeled data. Using traditional polls as labeled data can be a direction to address this challenge, but it also presents its own set of challenges. Polls can have errors and discrepancies, and there can be variations in the methodologies used by different pollsters. The timing of the polls can also pose challenges for traditional time-series approaches. To address these challenges, new techniques and approaches need to be developed. For example, one approach could be to combine traditional polling data with data from other sources, such as social media metrics, to improve the accuracy of predictions. Another approach could be to use unsupervised learning techniques to identify patterns and trends in the data that could help inform predictions. Additionally, there may be opportunities to develop new types of data collection methods that can provide more accurate and reliable data for training ML algorithms for election predictions.

B. Questions and Hypotheses for Research

This methodology seems to be a well-defined approach to answering the research questions and testing the hypotheses. By defining the media performance and using it to create an ML model trained with conventional elections, the study aims to predict election results and perform daily Direct casting with competitive results to conventional elections. Collecting data from various sources and running experiments using the proposed framework is a thorough approach to evaluating the effectiveness of the model. Additionally, comparing the

predictions with strong baselines provides a benchmark for evaluating the accuracy of the model's predictions. Overall, this methodology appears to be a robust approach to answering the research questions and testing the hypotheses.

RQ1
Is it possible to specify a procedure and develop an ML model that can forecast election outcomes based on candidate SM performance?

RQ2
Is it feasible to develop a procedure and build a machine learning model that can predict election results every day based on the SM performance of the candidates?To answer RQ1 and RQ2, this study considered a post positivist stance.

H1
Using an ML technique trained on traditional polls, it is feasible to establish a strategy and build a model based on the SM performance of candidates, which is capable of accurately forecasting.

H2
Using an ML method trained on traditional polls, it is feasible to establish a procedure and build a model based on the SM performance of candidates, which is capable of providing daily forecasts that are competitive with traditional polls.

Researchers seek to disprove the null hypotheses that result from these theories.

H1'
It is impossible to develop a procedure and build a model based on a candidate's SM performance using a machine learning technique trained on conventional polls, which is capable of accurately forecasting.

H2'
It is not possible to establish a procedure and develop a model based on the SM performance of candidates, using a machine learning technique trained using conventional polls, that is capable of providing election outcome forecasts with results that are competitive with traditional polls.

We created a technique with four basic phases in order to reject the null hypotheses of H1' and H2' in favour of the hypotheses H1 and H2, which provide solutions to the research questions: the concept of SM performance; the framework made up of a method and machine learning (ML) modeling for election prediction; the collection of all necessary data, including media, reviews, analysis made by psephologists.

THE SUGGESTED METHODOLOGY

A. Monitoring social media activity

The analysis of interactions on the credentials of candidates on various platforms like internet, Whatsapp, and Instagram is the primary method of measuring the performance of SM employed in this study. Opinions, reviews, and shares (or comparable phrases, such retweets on Whatsapp) on each post are included in the interactions that were analyzed. These behaviors show that the user has read the information, and responded to it. The measurement takes into account both the relative numbers per post and the absolute numbers of each interaction over the course of a time. These metrics are general and suitable for the majority of SM platforms that use newsfeeds. The performance evaluation is based on Zajonc's exposure theory, which postulates that merely subjecting a person to a stimuli repeatedly won't always result in learning. The more engagement a candidate receives on their SM posts, the more the SM algorithms will prioritize their content, leading to a snowball effect of more engagement and more exposure. Ultimately, this may lead to more positive attitudes towards the candidate and more votes. It's worth noting the same rationale of interactions may be applied to measure SM performance, even if the platforms present slight differences in their metrics.

B. MEF – A media framework for election.

Phase 1: Election Understanding In this phase, the goal is to understand the problem and define the objectives of the project. It involves understanding the context and the main stakeholders involved, as well as the goals and requirements of the project.

Phase 2: Data Collection and Understanding In this phase, the goal is to gather and explore the data necessary for the project. This includes collecting data from different sources, understanding the data quality and relevance, and performing exploratory data analysis to understand the distribution of data and identify any patterns or trends.

Phase 3: Data Preparation In this phase, the goal is to prepare the data for modeling. This includes removing the semantic errors, irrelevant data and choosing the appropriate patterns and features for modeling.

Phase 4: Modeling and Execution In this phase, the goal is to develop and execute the ML model. This involves selecting an appropriate algorithm, training and testing the model, and tuning the model to improve its performance.

Phase 5: Evaluation In this phase, the goal is to evaluate the performance of the model and assess its effectiveness in meeting the project objectives. This involves comparing the model's predictions with actual election results, identifying any errors or limitations of the model, and making recommendations for future improvements. The MEF process is designed to guide the development of an ML model for predicting election results based on social media performance. By following this process, researchers can ensure that their model is developed and evaluated in a systematic and rigorous way.

Fig. 1. The MEF Process.

C. Election Understanding

In addition to identifying the SM platforms and candidate profiles, it is also important to understand the characteristics and functionalities of each platform, as they may differ in terms of their user base, content format, and engagement metrics. For instance, while Whatsapp is known for its real-time updates and short-form text-based content, Instagram is more visual and favors images and videos. Facebook, on the other hand, offers a mix of both text and visual content and has a wider range of engagement options, such as reactions, comments, and shares. Therefore, understanding the nature of each platform and the type of content that resonates with its users is crucial for devising an effective social media strategy and for interpreting the SM performance data. Moreover, it is also important to gather information about the polling agencies and their methodology, as different pollsters may have different biases or sampling errors that can affect the accuracy of the predictions. It is recommended to select reputable pollsters with a track record of accurate predictions, and to obtain as much information as possible about their sampling methods, sample size, and margin of error. By doing so, one can have a better understanding of the limitations and uncertainties associated with the polling data, and adjust the modeling approach accordingly.

D. Gathering and Interpreting Data

Data from both SM and polls are collected in the second phase. In the data understanding phase, it is important to explore and analyze the collected data to gain insights into the data distribution, identify any anomalies or outliers, and determine the quality of the data. This process helps in determining which data preprocessing steps are required to prepare the data for the modeling phase. For instance, missing data may need to be imputed, and outliers may need to be removed or treated. In the case of SM data, it is also essential to extract relevant features from the raw data that can be used in the modeling phase. Examples of features that can be extracted from social media data include the number of likes, comments, shares, and followers of the candidates' social media profiles. With regards to poll data, it is important to verify the accuracy and reliability of the polls to be used. This is

because some polls may be biased or have a small sample size, leading to inaccurate results. Moreover, the time window within which polls are conducted is also critical as polling data closer to the Election Day is expected to be more accurate than those conducted earlier. It is also essential to understand the methodology used to conduct the polls and the target population for the polls. The data understanding phase is crucial in preparing the data for the modeling phase and ensuring that the predictions are accurate and reliable.

E. Data Preparation

It is important to perform data cleaning and preprocessing to ensure the accuracy and consistency of the data. This may involve removing duplicates, checking for missing values, and handling outliers. It is also important to normalize the data to ensure that different polls with different scales can be compared. Once the data is cleaned and preprocessed, feature engineering can be performed to create new features that may be more informative for the prediction model. For example, one could calculate the difference between the candidate's social media engagement and their opponents' engagement or create a sentiment analysis score for the candidate's posts. It is also important to split the data into training, validation, and testing sets to evaluate the performance of the model. In cases where the number of samples is small, cross-validation can be used to ensure that the model's performance is robust. Overall, the data preparation phase is crucial in ensuring that the data is suitable for the prediction model and can ultimately lead to more accurate predictions.

F. Modeling

Another important aspect to consider is the model's interpretability. This is particularly relevant in political prediction, where understanding the reasons behind the model's predictions is crucial. Models such as decision trees, random forests, and linear regression can provide insight into the contribution of each feature to the model's output, and therefore may be more interpretable. On the other hand, models such as neural networks and support vector machines (SVMs) can achieve high accuracy but may lack interpretability, especially in cases where the network is deep and complex. In these cases, additional techniques, such as saliency maps, may be used to identify the features that contribute most to the model's output. Another important consideration is the choice of architecture for the prediction model. One approach is to use an ensemble of models, which may improve the generalization performance of the system. For example, a committee machine can be constructed by training multiple regression models with different window sizes and then averaging their predictions. Convolutional neural networks (CNNs) or recurrent neural networks (RNNs) are used, that can capture complex temporal and sequential patterns in the SM data. However, these architectures may require large amounts of training data, which may not be available in the political prediction domain. The ML model should be chosen based on its ability to handle the small sample size, its generalizability, nonlinearity, interpretability, and the available computational resources. The choice of architecture should also take into account the data characteristics and the desired level of interpretability.

G. Evaluation

Forecasting and preprocessing is a big challenge in Elections. When evaluating the performance of ML models for election prediction, it is essential to consider the potential sources of bias in the data used for training and testing. For instance, the public polls used as a ground truth for evaluating the predictions may suffer from sampling biases, such as selection bias or non-response bias. Similarly, the social media data used as input to the prediction models may be biased towards certain demographic groups or political affiliations, which can impact the generalization ability of the model. Thus, it is important to carefully analyze the potential sources of bias and address them appropriately in the data preprocessing and model training stages to ensure reliable and unbiased predictions.

H. The MEF ML Model

The choices of MLP-BP and GRNN for the modelling technique and context of forecasting election results based on social media performance are well-justified and appear plausible. To combat overfitting and raise the model's accuracy, using PCA for dimensionality reduction and producing numerous datasets with different window sizes is also a viable strategy. The variance can be decreased and the final forecasts' accuracy can be improved by

employing a committee machine made up of an ensemble of 10 predictors. Both the MLP-BP and the GRNN are well-liked and often employed neural network designs with a track record of success in regression problems. The MLP-BP is a universal approximator, which means it is capable of approximating any continuous function, and has good generalization capabilities. The GRNN, on the other hand, is particularly suitable for small sample data, requires fewer hyper parameters, and is faster to train compared to the MLP-BP. Using both models in the ensemble can help to capture the strengths of each model and improve the accuracy of the predictions. The proposed modeling approach and the choice of MLP-BP and GRNN seem reasonable and well-justified for the context of predicting election results based on social media performance. However, it is important to evaluate the performance of the models thoroughly and compare them with well-defined baselines to determine their effectiveness.

I. MEF-DC: During the Campaign, a Media Framework for Elections

In the MEF-DC method, initial polls and SM performance data are gathered, the model is trained using the data that is now available, daily predictions are made using the SM data that has been collected, and the model is retrained using fresh labeled data when new poll data is published. It is crucial to choose the minimum number of polls required to train the model and start generating predictions because more labeled data results in more accurate forecasts. The MLP-BP and GRNN models are well suited for this situation, and retraining the model using fresh poll data is also crucial. It can be difficult to evaluate daily projections, therefore the most logical course of action is to do a descriptive, qualitative examination of the two candidates that received the most votes in terms of polls, predictions, and the final vote share. Additionally, prediction errors can be measured by comparing predictions to polls and considering competitive predictions as those that have errors below the historical MAE of 2.7 percentage points within a 3.00 deviation. This approach can also be seen as a cross-validation approach of the MEF.

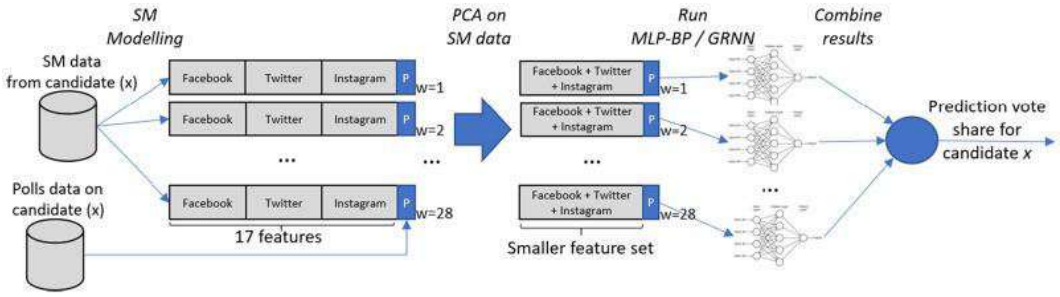

Fig. 2: MEF instantiation.

EXPERIMENTS

On recent Assembly elections in Himachal Pradesh, Gujarat, West Bengal, Tamil Nadu, and Karnataka, we conducted tests to anticipate the results. India's states all share a same background: they are geographically located in nearby proximity, have related roots, languages, traits, and electoral practices including voting, a large number of candidates running in the first round, and few polls are accessible. Applying the same technique to various states allows for a comparison of the outcomes and an evaluation of the framework.

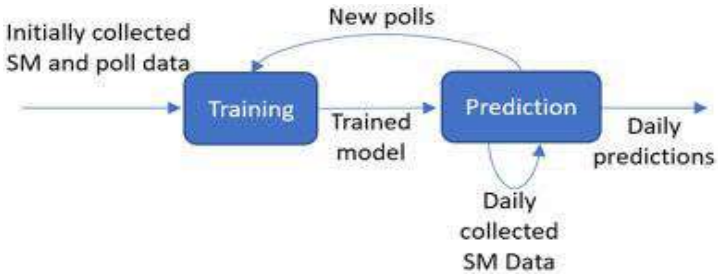

Fig. 3: The MEF-DC execution.

A. Modeling

The MLP-BP and GRNN were employed, as previously mentioned. Additionally, a basic linear regression model was applied. It seems that a lot of effort has been put into developing and optimizing the models used for the MEF-DC approach. The use of a committee machine for each model, composed of an ensemble of 10 predictors, is a good strategy to increase the robustness of the predictions. It is interesting to note that the MLP-BP model uses a fixed parameter approach with manually selected parameters, rather than a grid search parameter approach. This is a valid approach if the manually selected parameters are well-justified and have been shown to perform well in previous studies, as seems to be the case here. The use of the GRNN model is also a good choice, as it is well-suited to handling small sample sizes, which is likely the case in this context. It seems that the models used for the MEF-DC approach are well-suited to the task and have been carefully optimized to produce accurate predictions. Experiments were run twice, once with and once without PCA, to compare and validate the gain gained by using PCA. In all, six experiments were conducted: GRNN, GRNN with PCA, MLP-BP, GRNN with PCA, and linear regression with and without PCA.

DISCUSSION

It seems that the proposed approach has addressed several challenges that were present in previous studies on election prediction using social media data. The approach has addressed process challenges by defining the MEF and MEF-DC processes, and has addressed sampling challenges by using conventional polls to preprocess data by using data from different social media platforms. By individually training and forecasting the candidate results, the high vulnerability to volume manipulation was reduced. The study has also employed state-of-the-art machine learning models and has presented a clear strategy for parameter choice. Furthermore, the proposed approach has also addressed performance evaluation and scientific rigor challenges by using commonly used metrics, comparing predictions with end process and polls results, and performing statistical analyses of the results. Although the experiments were conducted in several states of India, the process and approach is simple enough to be employed in any country.

CONCLUSION

Overall, by suggesting a novel method for forecasting elections using electronic media data, this work has significantly advanced the disciplines of polls and electoral forecasts, electronic media studies, and Artificial intelligence. Process, sampling, modeling, and performance evaluation were some of the issues that the study addressed. It also offered a framework that may be used to apply to many elections with small alterations. In terms of machine learning, the work has applied ML to a novel and understudied environment that has particular difficulties since there is a dearth of historical data, the context quickly changes, there are few samples, and there is only one labeled variable—the outcome. The study looked at the prospect of discovering nonlinear connections between online behavior and offline results in social media studies, which is a challenging challenge for social scientists because of the large number of data required for these computations. Future research might expand on this study to look at how users interact with social media, how candidates utilize it as a platform for their campaigns, and how polarisation may impact elections globally. The suggested methodology has shown promise in real-time election prediction and might be a useful tool for election monitoring and analysis.

REFERENCES

[1] Abramowitz, A. I. (2016). Will time for change mean time for Trump? PS: Political Science & Politics, 49(4), 659–660.

[2] Armstrong, J. S. (1980). The Seer-Sucker theory: The value of experts in forecasting. Technology Review, 1980(83), 16–24.

[3] Biemer, P. P. (2010). Total survey error: Design, implementation, and evaluation. Public Opinion Quarterly, 74(5), 817–848.

[4] Campbell, J. E. (2016). Introduction. PS: Political Science & Politics, 49(4), 649–654.

[5] Machine Learning Approach to Predict the Accident Risk during Foggy Weather Conditions - Alliance International Conference on Artificial Intelligence and Machine Learning– Alliance University – Bangalore – April -2019

[6] Predicting election trends using data mining techniques - International Conference on future electric vehicular mobility and its challenges - JBREC - Hyderabad - 20th & 21st December 2021.

[7] Kalampokis E, Karamanou A, Tambouris E, Tarabanis KA. On predicting election results using Whatsapp and linked open data: the case of the UK 2010 election. J UCS. 2017;23(3):280–303.

[8] Sharma Y, Mangat V, Mandeep K. Sentiment analysis and opinion mining. Int J Soft Comput Artif Intell. 2015;3(1):59–62.

[9] Attarwala A, Dimitrov S, Obeidi A. How efficient is Whatsapp: Predicting 2012 US Presidential elections using Support Vector Machine via Whatsapp and comparing against Iowa Electronic Markets. In: Intelligent Systems Conference (IntelliSys), 7–8 September, London; 2017.

AUTHOR INFORMATION

M. Suresh Babu is a seasoned professional with over 26 years of experience in Teaching and Student Management. He is a Professor and Dean at Teegala Krishna Reddy Engineering College in Hyderabad. Given his various roles in the field, he has a strong background in computer science and engineering. Dr. M. Suresh Babu possesses a diverse skill set and has experience in both the software industry and academia. His experience spans both the technical and educational realms, showcasing a well-rounded professional with expertise in software development and a dedication to teaching and student management. He has a diverse and extensive involvement in different academic and administrative roles. Dr. M. Suresh Babu is actively involved in academic publications and holds various editorial roles in reputed journals. Additionally, he is a member of several professional bodies. He has contributed over 150 technical papers and is a freelance editor for several Telugu and English newspapers.

Dr. Pranayanath Reddy working as Associate Professor in the Department of CSE- at TKR Engineering College,Meerpet, Hyderabad . He did his Ph.D. in Cloud Computing from Pacific University, India & M.Tech from KITS , Warangal. His primary research interests include Information Security, Cloud computing, Software Engineering and Software Testing. He has 14 Years of Teaching Experience and 2 years of IT Industry. He worked at Infosys Technologies , as a Senior System Engineer. He worked with top Deemed and Private Universities at Hyderabad. He has published 20+ International papers in journals and conferences. He also published 3 patents. He is certified in Python Programming conducted by Microsoft. He is a highly versatile teacher in various Computer related subjects such as Cloud computing, database management system, Web Technologies, IoT, Software Engineering and many other subjects related to Computer Science and Engineering.

He has presented papers in National and International forums, published number of articles in journals having high impact factors, and chaired many conferences.

Dharmavaram Asha Devi, Member IEEE, Fellow IETE. The Author is working as a Professor in the Department of Electronics & Communication Engineering, SNIST, Hyderabad. She became an Associate Member of IETE in Dec-2000 from the Institute of Electronics and Telecommunication Engineering, New Delhi, India. She awarded M.Tech. Degree in Digital Systems and Computer Electronics from JNTU Anantapur in 2005. She achieved a Doctorate Degree from Sri Krishnadevaraya University, Anantapur in Dec 2011, for her research contribution in Embedded System based Soil Analysis under the supervision of Prof. K. Malakondaiah, from SKU, Anantapur. She has published more than 35 papers in International Journals and presented 40 plus research papers in International level conferences on behalf of her research work done in Embedded Systems and VLSI Designs. Her present research area is FPGA based embedded applications, design and verification of SoC designs using both front end and back end VLSI CAD tools.

B. Deevena Raju holds a diverse range of educational qualifications and has amassed significant experience in the field of CSE. Dr Deevena Raju is a seasoned professional with over 19 years of experience in Teaching and Student Management. He is an Assistant Professor in Faculty of Science & Technology – ICFAI Tech School,n Hyderabad. Given his various roles in the field, he has a strong background in computer science and engineering. He has provided guidance and mentorship to both M.Tech and B.Tech students, facilitating the successful completion of their projects and aiding many students in obtaining their degrees. Additionally, assisted students in securing teaching positions as well as employment opportunities in software companies, thereby contributing to their professional growth and success. He worked with top Deemed and Private Universities at Hyderabad and published 30+ papers in reputed journals and conferences