Personalized diabetes diagnosis using machine learning and electronic health records

Gowthami S.[1], R. Venkata Siva Reddy[1], Mohammed Riyaz Ahmed[2]
[1]School of Electronics and Communication Engineering, REVA University, Bangalore, India
[2]Department of Electronics and Communication Engineering, HKBK College of Engineering, Bangalore, India

## Article Info

ABSTRACT

Diabetes mellitus (DM) poses a significant health challenge globally, necessitating accurate and timely diagnosis for effective management. Conventional diagnostic methods often struggle to address the multifaceted nature of diabetes and the requisite lifestyle adjustments. In this study, we propose a data-driven approach utilizing machine learning techniques to enhance diabetes diagnosis. By leveraging extensive patient attributes and medical records, machine learning algorithms can uncover intricate patterns and correlations. Our methodology, validated on the PIMA India dataset, demonstrates promising results. The random forest model achieved the highest accuracy of 87%, followed closely by gradient boost at 90%. Notably, XGBoost and CATBoost models attained a peak accuracy of 90.9%. These findings underscore the potential of machine learning in transforming diabetes diagnosis. Beyond improving diagnostic accuracy, our approach aims to guide individuals towards healthier lifestyles. Intelligent systems driven by machine learning hold promise for revolutionizing diabetes management, ultimately leading to better patient outcomes and more effective health care delivery.
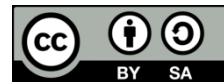
Corresponding Author:

Gowthami S
School of Electronics and Communication Engineering, REVA University
Bangalore-560064, India
Email: sgowthami12@gmail.com

1. INTRODUCTION

Diabetes mellitus (DM) poses a substantial global health challenge, with its prevalence escalating due to factors such as sedentary lifestyles, dietary changes, and demographic shifts towards an aging population. According to the World Health Organization (WHO), the number of people with diabetes has risen from 108 million in 1980 to 422 million in 2014, with projections indicating a further increase to 642 million by 2040 [1]. This exponential growth underscores the urgent need for effective management strategies. The complexity of DM lies in its multifaceted nature, characterized by variations in etiology, presentation, and response to treatment. Type 2 diabetes, the most common form, is particularly influenced by lifestyle factors such as diet and physical activity, contributing to its increasing prevalence. Timely and accurate diagnosis is paramount for initiating appropriate management strategies and preventing long-term complications associated with uncontrolled diabetes [2]. Complications include cardiovascular disease, kidney failure, blindness, and lower limb amputation. Therefore, proactive measures, including early screening and intervention, are crucial in mitigatingthe impact of diabetes on individuals and healthcare systems globally.

Traditional diagnostic methods for diabetes, such as fasting blood glucose tests and oral glucose tolerance tests, have long been relied upon for initial screening and diagnosis. However, these methods possess inherent limitations that hinder their ability to comprehensively capture the diverse manifestations of

the disease. For instance, fasting blood glucose tests may overlook subtle fluctuations in blood glucose levels that occur throughout the day, potentially leading to missed diagnoses or delayed intervention [3]. Similarly, oral glucose tolerance tests, while useful in certain scenarios, may not adequately account for individual variations in metabolic responses, resulting in inaccuracies in diagnosis or classification of diabetes subtypes. The manual analysis of extensive electronic health records (EHRs) further compounds the challenges faced by healthcare providers in accurately diagnosing and managing diabetes [4]. EHRs contain a wealth of patient data, ranging from medical history and laboratory results to clinical notes and imaging studies. However, the sheer volume and complexity of this information make it difficult for healthcare professionals to efficiently extract relevant insights and make informed decisions. Additionally, discrepancies or inaccuracies in EHR documentation can further impede the diagnostic process, leading to potential errors or delays in treatment initiation.

Figure 1 shows that diabetes can be diagnosed using either the criteria of hemoglobin A1c or plasma glucose concentration, with a fasting plasma glucose (FPG) level exceeding 126 mg/dL (7.0 mmol/L) indicating a positive diagnosis [5], [6]. The two-hour oral glucose tolerance test (OGTT) is an alternative method, requiring plasma glucose level measurement before and two hours after consuming 75 grams of glucose. If the postprandial plasma glucose (PG) level exceeds 200 mg/dL (11.1 mmol/L), it confirms diabetes mellitus (DM) [7]. Given its inconvenience and cost, OGTT is less favored compared to FPG. Managing diabetes, with its long-term health implications, requires vigilant monitoring of blood glucose levels to prevent and address complications. While living with diabetes is challenging, appropriate support and care make it manageable. Intelligent diabetes diagnosis systems, inspired by biological constructs, offer efficient tools for management [8]. Machine learning techniques have notably advanced diabetes diagnosis, aiding healthcare professionals by analyzing diverse patient attributes and medical records to unveil hidden patterns and facilitate timely identification of diabetes.

As the global prevalence of diabetes rises and healthcare systems strive for more personalized and efficient care, there is increasing interest in leveraging advanced technologies such as machine learning (ML) to complement traditional diagnostic methods. ML algorithms offer a powerful tool for analyzing vast datasets, including electronic health records (EHRs), to uncover intricate patterns, correlations, and risk factors associated with diabetes [9], [10]. By harnessing ML, it becomes feasible to develop sophisticated diagnostic models tailored to individual patient profiles, thereby enhancing both the precision and timeliness of diabetes diagnosis and management. These ML-driven models have the capacity to sift through extensive patient data, identifying subtle indicators and predictive markers that may elude conventional approaches [11], [12]. This enables healthcare professionals to make more informed decisions regarding diagnosis, treatment strategies, and ongoing patient monitoring, leading to improved health outcomes. Furthermore, the dynamic nature of ML allows for continual refinement of diagnostic models, ensuring adaptability and effectiveness over time. As ML technologies advance, their integration into healthcare systems holds the potential to revolutionize diabetes management, offering proactive and personalized care that ultimately benefits individuals affected by this pervasive chronic condition [13].
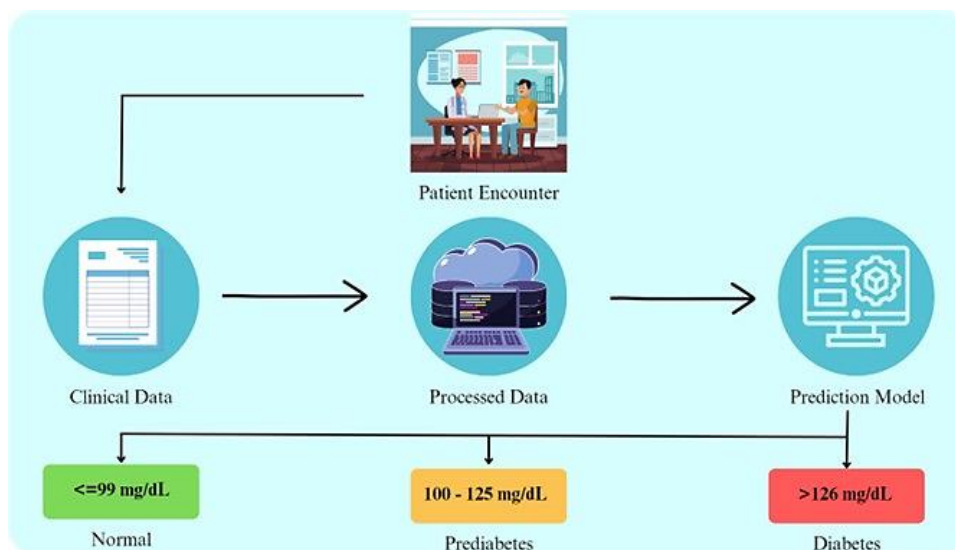

Figure 1. Block diagram representation of diagnosis of diabetes

2. BACKGROUND

The evolution of diabetes diagnosis and prediction technologies as shown in Figure 2, particularly with the integration of machine learning (ML), has been remarkable since the 1990s. Initially, diagnosis relied heavily on traditional methods such as fasting blood glucose tests and oral glucose tolerance tests, which provided valuable but limited insights into glycemic status. However, advancements in technology, coupled with a deeper understanding of the disease, have led to significant improvements in diagnostic accuracy and predictive capabilities.

In the 1990s and early 2000s, the advent of continuous glucose monitoring (CGM) systems revolutionized diabetes management by enabling real-time monitoring of glucose levels [14]. Unlike traditional finger stick glucose testing, CGM systems offered individuals with diabetes and healthcare providers a continuous stream of data, providing insights into glucose trends, variability, and patterns throughout the day and night. This continuous monitoring capability allowed for more proactive adjustments to insulin therapy, diet, and lifestyle, leading to improved glycemic control and reduced risk of hypoglycemia and hyperglycemia-related complications. The introduction of hemoglobin A1c (HbA1c) testing during this time provided a measure of long-term glycemic control, reflecting average blood glucose levels over the preceding 2-3 months [15], [16]. This standardized metric became a cornerstone in diabetes management, aiding in the diagnosis, treatment adjustment, and long-term monitoring of diabetes. Together, CGM and HbA1c testing represented significant advancements in diabetes care, empowering individuals with diabetes to take a more active role in managing their condition and facilitating more personalized and targeted treatment strategies by healthcare providers.

The integration of machine learning and data analytics into diabetes diagnosis and prediction emerged as a significant advancement in the 2010s [17], [18]. ML algorithms began to analyze large datasets, including electronic health records (EHRs), genomic data, and continuous glucose monitoring data, to identify predictive biomarkers and risk factors associated with diabetes [19], [20]. These algorithms could uncover intricate patterns and correlations within the data, enabling more accurate diagnosis and prediction of diabetes onset. Moreover, ML-driven models facilitated personalized treatment strategies tailored to individual patient profiles, improving overall patient outcomes. The evolution of telemedicine platforms and remote monitoring technologies has further enhanced diabetes diagnosis and prediction [21], [22]. Telemedicine allows for remote assessment of patients' glycemic status and enables healthcare providers to intervene promptly in cases of abnormal glucose levels, facilitating early diagnosis and management of diabetes. Additionally, ongoing research endeavors continue to focus on biomarker discovery and the development of novel diagnostic tools, further advancing the field of diabetes diagnosis and prediction [23].
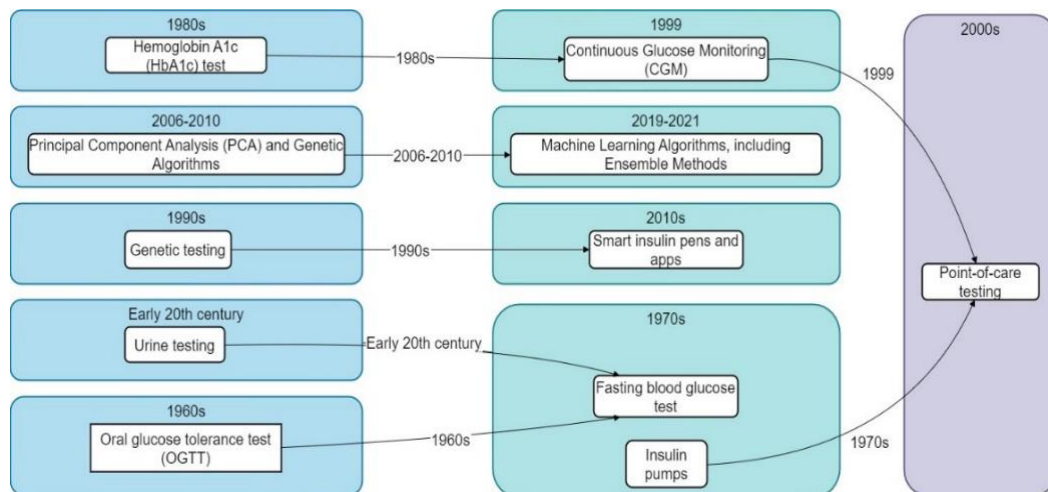

Figure 2. Block diagram representation of diagnosis of diabetes

3. METHOD

The first step involved obtaining a comprehensive dataset of anonymized patient records, including demographic details, medical history, and laboratory results relevant to diabetes diagnosis. Data were sourced from reputable healthcare databases, ensuring quality and compliance with privacy regulations. The dataset's representativeness and diversity were carefully considered, and patient information was anonymized to protect privacy. This dataset served as the basis for subsequent model development and evaluation stages.

The PIMA Indian dataset utilized in this study provides a valuable resource for exploring and developing diabetes diagnosis models [24]. The dataset comprises comprehensive information collected from 768 PIMA Indian women, encompassing various features crucial for understanding and predicting diabetes out- comes. Figures 3 and 4 convey the analysis of various features such as glucose levels, blood pressure, body mass index (BMI), skin thickness, insulin levels, age, and whether or not the individual has diabetes. Figure 3 shows the exploratory data analysis for categorical attributes: Figure 3(a) illustrates the age distribution among the individuals, while Figure 3(b) depicts the BMI categories and their distribution. Figure 3(c) presents the distribution of individuals with and without children, and Figure 3(d) highlights the glucose level categories. Furthermore, Figure 3(e) focuses on the distribution of hypoglycemic status, Figure 3(f) on obesity status, and Figure 3(g) on pregnancy status among the participants.

(a)
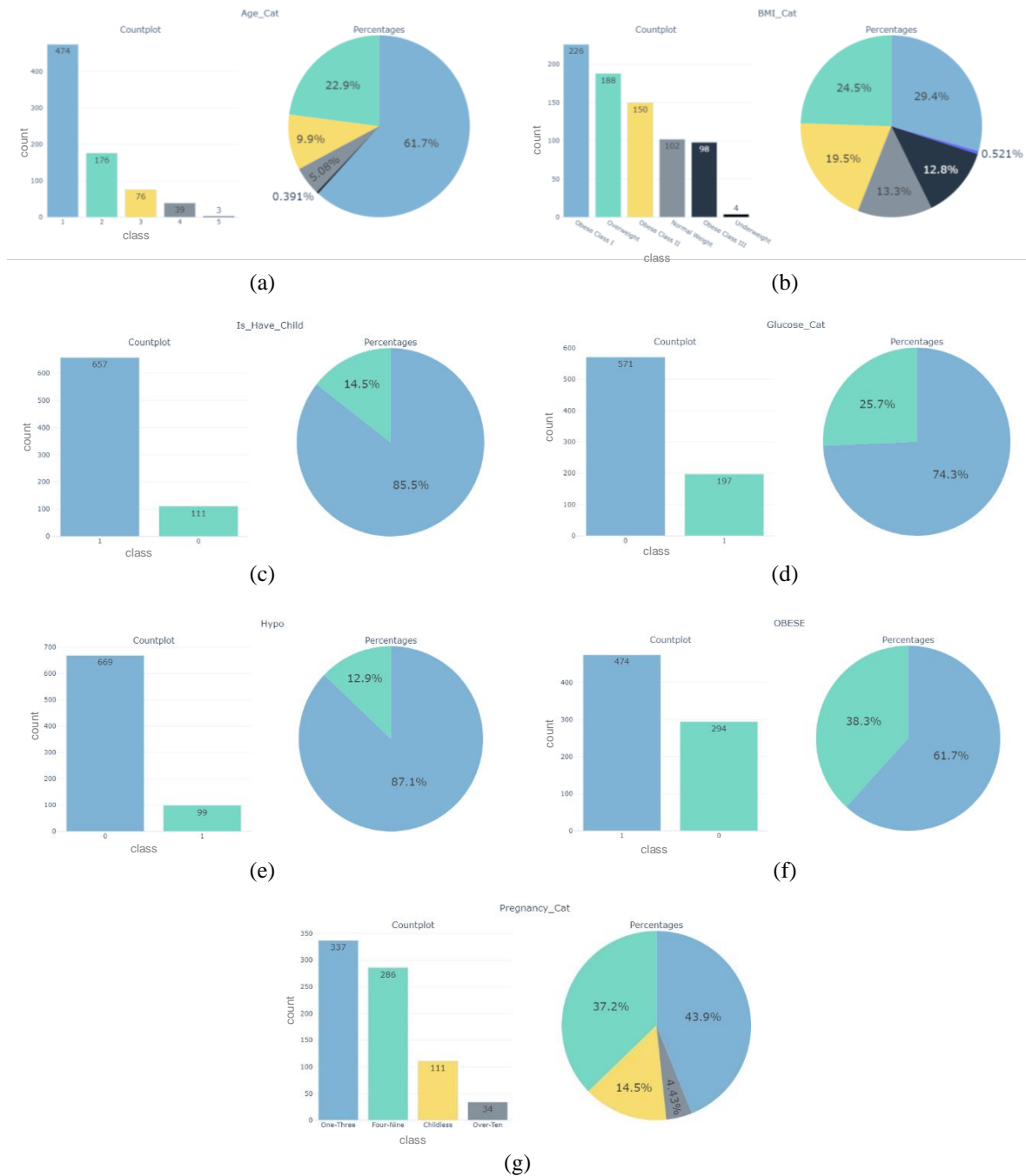

(b)

(c)

(d)

(e)

(f)

(g)

Figure 3. Exploratory data analysis for categorical attributes: (a) age, (b) BMI, (c) having a child, (d) glucose, (e) hypo, (f) obese, and (g) pregnancy

In contrast, Figure 4 presents the numerical exploratory data analysis on various attributes, offering a more detailed quantitative assessment: Figure 4(a) shows the distribution of ages in the dataset, and Figure 4(b) details blood pressure measurements across the population. Figure 4(c) provides a detailed analysis of BMI values, while Figure 4(d) covers diastolic pressure gradient (DPG) statistics. Additionally, Figure 4(e) displays glucose levels in numerical terms, Figure 4(f) analyzes the number of pregnancies, and Figure 4(g) presents skin thickness measurements. By analyzing both categorical and numerical attributes, these figures collectively enhance the understanding of the dataset, facilitating the development of effective diagnostic models for diabetes.
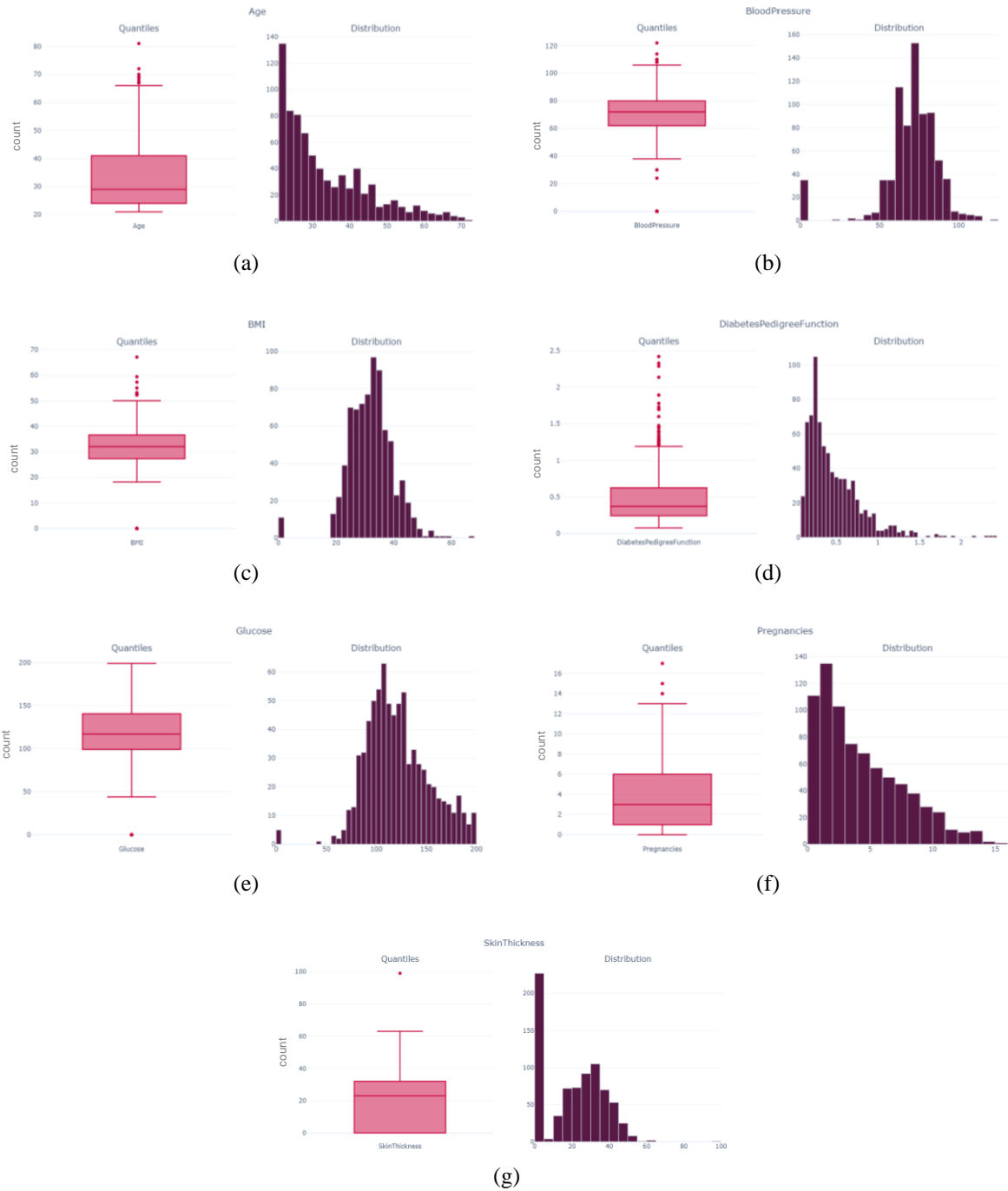

Figure 4. Numerical exploratory data analysis on attributes: (a) age, (b) blood pressure, (c) BMI, (d) DPG, (e) glucose, (f) pregnancies, (g) skin thickness

With its diverse features, the PIMA Indian dataset enables researchers to investigate the intricate relationships between various health indicators and the likelihood of diabetes. For example, glucose levels, blood pressure, and BMI are well-known risk factors for diabetes, while skin thickness and insulin levels provide additional insights into the disease's physiological mechanisms [25]. The dataset also includes age information, which is essential for understanding the impact of aging on diabetes prevalence. As shown in Figure 3, by analyzing this dataset, we gained valuable insights into the complex interplay of these factors and helped in developing effective models for diagnosing diabetes in not only the PIMA Indian population but potentially other populations as well. Data preparation plays a crucial role in machine learning projects, and in this study, several preprocessing tasks were performed to ensure the data's quality and suitability for analysis. The first step involved handling missing values in the dataset. Missing values can introduce biases and affect the accuracy of the models. Various techniques, such as imputation or removing rows with missing values, can be employed to address this issue. By carefully addressing missing values, we ensured that the dataset used for analysis was complete and reliable. Another important pre-processing step involved normalizing features. Normalization is carried out to ensure that various features are brought to a comparable scale, preventing any individual feature from exerting excessive influence over the learning process. This step helps optimize model performance and can be particularly useful when dealing with features with different ranges or units. Additionally, categorical variables were encoded using techniques like one-hot encoding, which converts categorical features into binary vectors. This enables the machine learning models to understand and utilize the information contained in categorical variables effectively.

Once preprocessing was completed, the dataset was divided into two subsets: training and testing. The training subset was employed to train the machine learning models, while the testing subset acted as a separate dataset to assess the models' performance. This separation guarantees that the models are evaluated on independent data and prevents bias from evaluating their performance on the data they were trained on. Furthermore, by evaluating the models on unseen data, we can assess their generalization ability and ensure that they perform well on new, unseen instances. Overall, these preprocessing tasks, including handling missing values, normalizing features, and splitting the dataset, were performed to ensure that the data used for analysis was high quality and suitable for developing accurate and reliable machine learning models. Combined with appropriate feature engineering techniques, these steps help optimize the models' performance and enhance their ability to make accurate predictions and generate meaningful insights from the data feature selection is a vital step in building effective and interpretable models, and in this study, we utilized various techniques to identify the most informative features. One approach involved correlation analysis, which measures the relationship between features and the target variable. By analyzing the correlation coefficients, we could identify the features strongly correlated with the diabetes outcome. This allowed us to focus on the most relevant variables and exclude any redundant or irrelevant ones from our model. Additionally, we employed mutual information, a statistical measure that quantifies the amount of information shared between variables. This technique helped us identify features that provided significant information about the diabetes outcome, even if they did not exhibit strong linear relationships. By considering mutual information, we were able to capture non-linear dependencies and uncover potentially valuable features that may have been overlooked through correlation analysis alone. Furthermore, we utilized recursive feature elimination, a technique that recursively eliminates features based on their importance to the model. This method enabled us to assess the impact of each feature on the model's performance and select the subset of features that contributed most significantly to accurate predictions. By reducing the number of features, we aimed to improve model simplicity and generalization and enhance the interpretability of our diabetes diagnosis model. Through these feature selection techniques and the analysis of corresponding graphs, we identified the most informative features for our diabetes diagnosis model. This allowed us to build a more accurate and interpretable model by focusing on the relevant variables and eliminating unnecessary complexity.

In constructing our diabetes diagnosis model, we utilized diverse machine learning algorithms, each offering unique advantages in addressing the diagnostic task. Logistic regression is a widely-used algorithm that provides interpretable results and can effectively model the relationship between the features and the probability of diabetes. Random forest is an ensemble algorithm combining numerous decision trees to construct a strong and precise model. Its capability to handle non-linear relationships and interactions among features makes it well-suited for intricate datasets, such as the PIMA Indian dataset. We also employed gradient boosting algorithms, including CatBoost, XGBoost, GradientBoost, and the LGBMClassifier. Ensembled based algorithms excel at capturing complex patterns and handling high-dimensional data, making them well-suited for our diabetes diagnosis task. We utilized multiple algorithms to compare their performances and identify the most effective approach for accurately predicting diabetes outcomes. This approach allowed us to leverage the strengths of each algorithm and select the model that exhibited the best balance of accuracy, interpretability, and generalizability.

Evaluating the developed models is crucial in understanding their performance and reliability in diabetes diagnosis. To assess the diagnostic capabilities of our models, we employed various metrics. Accuracy is a widely employed metric that evaluates the overall accuracy of the model's predictions. It is calculated by dividing the number of correct predictions by the total number of predictions. Accuracy provides a comprehensive evaluation of the model's performance; however, it may not be appropriate when working with imbalanced datasets. Precision is a metric that estimates the percentage of accurately predicted positive instances out of all the positive predictions made by the model. It is calculated by dividing the number of true positives by the sum of true and false positives. Precision indicates how well the model avoids false positive predictions, which is important in preventing unnecessary treatments or interventions. Recall, also called sensitivity or true positive rate, quantifies the proportion of correctly predicted positive instances out of all the positive cases. It is calculated by dividing the number of true positives by the sum of true positives and false negatives. Recall plays a crucial role in identifying all the true positive cases, ensuring that the model does not overlook any potential instances of diabetes. The F1-score is a metric that combines precision and recall in a balanced way by using the harmonic mean, ensuring an equitable measure of both aspects. It is calculated as The F1-score combines precision and recall into a single value, which is useful when there is a trade-off between precision and recall. The receiver operating characteristic (ROC) curve analysis is also employed to evaluate the model's performance across different probability thresholds. The ROC curve graphically represents the relationship between the true positive rate (sensitivity) and the false positive rate (1-specificity) at different threshold settings. The area under the ROC curve (AUC-ROC) is a widely used metric to summarize the performance of the model. A higher AUC-ROC indicates that the model has a better ability to distinguish between classes, showcasing its effectiveness in making accurate predictions. By analyzing these metrics and performing cross-validation to validate our model's generalization, we could assess their performance comprehensively. Finally, the results obtained from this evaluation were compared, enabling us to identify the most promising model for accurate diabetes diagnosis.

4. RESULT AND DISCUSSION

In this study, our primary objective was to develop an effective diabetes diagnosis model using a variety of machine learning algorithms applied to the PIMA India dataset. To ensure the reliability and accuracy of our model, we conducted thorough preprocessing steps, which included descriptive analysis and feature selection. These steps allowed us to identify and incorporate the most informative variables into our model, thereby enhancing its predictive power. Our evaluation of the model's performance was based on several key metrics, including accuracy, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC), which collectively provide a comprehensive assessment of the model's ability to correctly classify individuals with and without diabetes.

Our findings revealed notable differences in performance among the various machine learning algorithms tested. Figures 5(a) to (e) presents a comparison of various machine learning models based on their performance metrics. The CatBoost classifier, illustrated in Figure 5(a), had an accuracy of 96.5%, a precision of 86%, but a recall of 0.0%. The decision tree model in Figure 5(b) achieved an accuracy of 84.3%, a precision of 78.1%, and a recall of 84.4%. The gradient boost model in Figure 5(c) achieved an accuracy of 95.4%, a precision of 85.7%, and a recall of 90.3%. The lightGBM classifier, shown in Figure 5(d), recorded an accuracy of 95.6%, a precision of 83.5%, and a recall of 89.1%. The logistic regression model achieved an accuracy of 87.4%, a precision of 67.3%, and a recall of 77.9% in Figure 5(e). The random forest model, depicted in Figure 5(f), demonstrated an accuracy of 93.1%, a precision of 80.3%, and a recall of 87%. Lastly, the XGBoost classifier in Figure 5(g) achieved an accuracy of 95.9%, a precision of 86.8%, and a recall of 90.9%. Overall, while the decision tree and logistic regression models showed respectable accuracies, the CatBoost classifier stood out as the top performer with the highest accuracy of 96.5%, despite its recall of 0.0%. Furthermore, the gradient boost model surpassed even this high benchmark, achieving an accuracy of 90% as shown in Table 1. The consistently high accuracies observed across other advanced algorithms such as LGBMClassifier, XGBoost, and CATBoost, further underscored the robustness and effectiveness of ensemble methods in diabetes diagnosis.

Comparing our results with previous studies, our research contributes to the growing body of evidence supporting the efficacy of machine learning in diabetes diagnosis. Notably, our study highlights the potential for ensemble methods to significantly enhance diagnostic accuracy, offering promising prospects for improving patient outcomes through early detection and intervention. However, it is essential to acknowledge certain limitations, such as the relatively small size of the dataset and potential biases inherent in the PIMA India dataset, which may affect the generalizability of our findings.

Moving forward, future research endeavors could delve deeper into exploring the underlying factors influencing model performance and investigating novel approaches to further improve diagnostic accuracy. This may involve exploring advanced feature engineering techniques, leveraging domain knowledge-driven features, or integrating data from external sources to enrich the dataset. By addressing these considerations and continuing to refine diagnostic models, we can advance the field of diabetes diagnosis and ultimately contribute to better healthcare outcomes for individuals affected by this chronic condition.
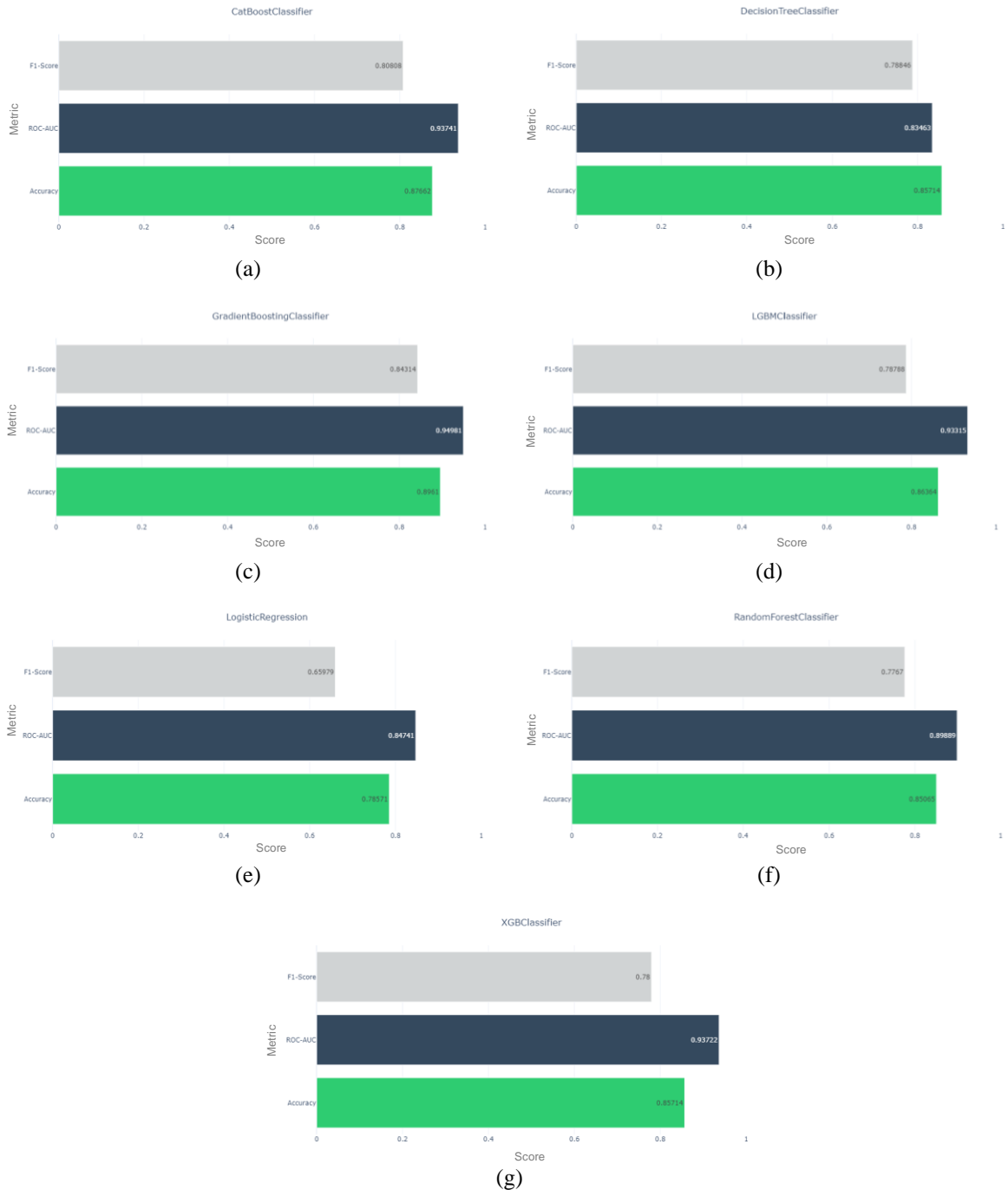
Figure 5. Graphical representation of F1-score, ROC-AUC and Accuracy for (a) CatBoost classifier, (b) decision tree classifier, (c) gradient boost classifier, (d) light gradient boost machine classifier, (e) logistic regression, (f) random forest classifier, and (g) extreme gradient boost classifier

Table 1. Metrics of ML algorithms after being trained with the proposed model

SL No.	Algorithms	ROC-AUC	F1-score	Accuracy
1	Logistic regression	87.4%	67.3%	77.9%
2	Random forest	93.1%	80.3%	87%
3	Gradient boost	95.4%	85.7%	90.3%
4	Lgbm classifier	95.6%	83.5%	89.1%
5	Xgb classifier	95.9%	86.8%	90.9%
6	Catboost classifier	96.5%	86%	0.0%
7	Decision tree	84.3%	78.1%	84.4%

5.　CONCLUSION

Integrating machine learning techniques in diabetes diagnosis presents a promising avenue for improving healthcare outcomes in the face of the growing diabetes epidemic. By leveraging the power of data analysis and advanced algorithms, machine learning enables accurate and timely detection of diabetes, empowering healthcare professionals to provide personalized care and interventions. In addition, the comprehensive analysis of patient attributes and medical records allows for a holistic understanding of the disease, enabling early detection and prediction. This research underscores the significance of addressing the challenges posed by diabetes and emphasizes the role of machine learning in transforming diagnosis practices. The development of intelligent systems that harness the potential of machine learning holds immense potential for improving the lives of individuals with diabetes. By enhancing diagnostic accuracy, these systems can guide healthcare professionals in making informed decisions and tailoring treatment strategies to meet each patient's specific needs. Furthermore, the application of machine learning techniques in diabetes diagnosis aligns with the evolving landscape of healthcare, where data-driven approaches are becoming increasingly prevalent. Yet, it is important to acknowledge the limitations and challenges associated with machine learning in diabetes diagnosis. Ethical considerations, such as data privacy and security, must be carefully addressed to ensure the responsible use of patient data. Additionally, the interpretability and transparency of machine learning algorithms require ongoing research and development to build trust and confidence among healthcare professionals and patients. Moving forward, further research and collaboration between healthcare providers, data scientists, and policy-makers are essential to realizing the potential of machine learning in diabetes diagnosis fully. We can create robust systems that seamlessly integrate into clinical practice and improve patient outcomes by addressing the existing challenges and refining the algorithms and models.

REFERENCES

[1]　K. L. Ong *et al.*, "Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021," *The Lancet*, vol. 402, no. 10397, pp. 203–234, Jul. 2023, doi: 10.1016/S0140-6736(23)01301-6.

[2]　E. U. Alejandro *et al.*, "Gestational diabetes mellitus: a harbinger of the vicious cycle of diabetes," *International Journal of Molecular Sciences*, vol. 21, no. 14, Jul. 2020, doi: 10.3390/ijms21145003.

[3]　M. A. Nauck, D. R. Quast, J. Wefers, and J. J. Meier, "GLP-1 receptor agonists in the treatment of type 2 diabetes – state-of-the-art," *Molecular Metabolism*, vol. 46, Apr. 2021, doi: 10.1016/j.molmet.2020.101102.

[4]　T. Zhu, K. Li, P. Herrero, and P. Georgiou, "Deep learning for diabetes: a systematic review," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 7, pp. 2744–2757, Jul. 2021, doi: 10.1109/JBHI.2020.3040225.

[5]　W. Hu *et al.*, "High L-valine concentrations associate with increased oxidative stress and newly-diagnosed type 2 diabetes mellitus: a cross-sectional study," *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*, vol. 15, pp. 499–509, Feb. 2022, doi: 10.2147/DMSO.S336736.

[6]　R. Karnchanasorn *et al.*, "Comparison of the current diagnostic criterion of HbA1c with fasting and 2-hour plasma glucose concentration," *Journal of Diabetes Research*, vol. 2016, pp. 1–11, 2016, doi: 10.1155/2016/6195494.

[7]　V. Ahuja *et al.*, "Accuracy of 1-hour plasma glucose during the oral glucose tolerance test in diagnosis of type 2 diabetes in adults: a meta-analysis," *Diabetes Care*, vol. 44, no. 4, pp. 1062–1069, Apr. 2021, doi: 10.2337/dc20-1688.

[8]　D. Lee and S. N. Yoon, "Application of artificial intelligence-based technologies in the healthcare industry: opportunities and challenges," *International Journal of Environmental Research and Public Health*, vol. 18, no. 1, Jan. 2021, doi: 10.3390/ijerph18010271.

[9]　J. Guru, S. R, G. H L, V. Ravi, M. Almeshari, and Y. Alzamil, "Electronic health record (EHR) system development for study on EHR data-based early prediction of diabetes using machine learning algorithms," *The Open Bioinformatics Journal*, vol. 16, no. 1, Oct. 2023, doi: 10.2174/18750362-v16-e230906-2023-15.

[10]　S. B. Junaid *et al.*, "Recent advancements in emerging technologies for healthcare management systems: a survey," *Healthcare*, vol. 10, no. 10, p. 1940, Oct. 2022, doi: 10.3390/healthcare10101940.

[11]　M. Elbadawi *et al.*, "Disrupting 3D printing of medicines with machine learning," *Trends in Pharmacological Sciences*, vol. 42, no. 9, pp. 745–757, Sep. 2021, doi: 10.1016/j.tips.2021.06.002.

[12] S. Rasool, A. Husnain, A. Saeed, A. Y. Gill, and H. K. Hussain, "Harnessing predictive power: exploring the crucial role of machine learning in early disease detection," *JURIHUM: Jurnal Riset dan Humaniora*, vol. 1, no. 2, pp. 302–315, 2023.

[13] Z. Ahmed, K. Mohamed, S. Zeeshan, and X. Dong, "Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine," *Database*, vol. 2020, Jan. 2020, doi: 10.1093/database/baaa010.

[14] H. Teymourian, A. Barfidokht, and J. Wang, "Electrochemical glucose sensors in diabetes management: an updated review (2010–2020)," *Chemical Society Reviews*, vol. 49, no. 21, pp. 7671–7709, 2020, doi: 10.1039/D0CS00304B.

[15] K. J. Kim *et al.*, "Time to reach target glycosylated hemoglobin is associated with long-term durable glycemic control and risk of diabetic complications in patients with newly diag- nosed type 2 diabetes mellitus: a 6-year observational study," *Diabetes & Metabolism Journal*, vol. 45, no. 3, pp. 368–378, May 2021, doi: 10.4093/dmj.2020.0046.

[16] M. S. Joy, W. T. Cefalu, S. L. Hogan, and P. H. Nachman, "Long-term glycemic control measurements in diabetic patients receiving hemodialysis," *American Journal of Kidney Diseases*, vol. 39, no. 2, pp. 297–307, Feb. 2002, doi: 10.1053/ajkd.2002.30549.

[17] S. Benjamens, P. Dhunnoo, and B. Meskó, "The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database," *npj Digital Medicine*, vol. 3, no. 1, Sep. 2020, doi: 10.1038/s41746-020-00324-0.

[18] D. Tsay and C. Patterson, "From machine learning to artificial intelligence applications in cardiac care," *Circulation*, vol. 138, no. 22, pp. 2569–2575, Nov. 2018, doi: 10.1161/CIRCULATIONAHA.118.031734.

[19] M. Bernardini, M. Morettini, L. Romeo, E. Frontoni, and L. Burattini, "Early temporal prediction of type 2 diabetes risk condition from a general practitioner electronic health record: a multiple instance boosting approach," *Artificial Intelligence in Medicine*, vol. 105, May 2020, doi: 10.1016/j.artmed.2020.101847.

[20] M. Vettoretti, G. Cappon, G. Acciaroli, A. Facchinetti, and G. Sparacino, "Continuous glucose monitoring: current use in diabetes management and possible future applications," *Journal of Diabetes Science and Technology*, vol. 12, no. 5, pp. 1064–1071, Sep. 2018, doi: 10.1177/1932296818774078.

[21] J. Golledge, M. Fernando, P. Lazzarini, B. Najafi, and D. G. Armstrong, "The potential role of sensors, wearables and telehealth in the remote management of diabetes-related foot disease," *Sensors*, vol. 20, no. 16, Aug. 2020, doi: 10.3390/s20164527.

[22] S. Ellahham, "Artificial intelligence: the future for diabetes care," *The American Journal of Medicine*, vol. 133, no. 8, pp. 895–900, Aug. 2020, doi: 10.1016/j.amjmed.2020.03.033.

[23] J. M. Yoshizawa, C. A. Schafer, J. J. Schafer, J. J. Farrell, B. J. Paster, and D. T. W. Wong, "Salivary biomarkers: toward future clinical and diagnostic utilities," *Clinical Microbiology Reviews*, vol. 26, no. 4, pp. 781–791, Oct. 2013, doi: 10.1128/CMR.00021-13.

[24] C. C. Olisah, L. Smith, and M. Smith, "Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective," *Computer Methods and Programs in Biomedicine*, vol. 220, pp. 1–12, Jun. 2022, doi: 10.1016/j.cmpb.2022.106773.

[25] M. Napolitano, M. Megna, and G. Monfrecola, "Insulin resistance and skin diseases," *The Scientific World Journal*, vol. 2015, pp. 1–11, 2015, doi: 10.1155/2015/479354.

BIOGRAPHIES OF AUTHORS

Gowthami S. 🆔 🔍 SC ◉ (Registered and practicing patent Attorney) received the B.E. and M.Tech degrees in information science and engineering and software engineering from Visvesvaraya Technological University Belgaum, India, in 2013 and 2015, respectively. She started her career with a start-up company Elegant Aero Sports, a Training Institute to teach students, flying models and robotics. As she aspired to become an entrepreneur, she founded Allinnov R&D Pvt. Ltd in 2015 as a Startup Company headquartered in Mysuru, along with Mr. Ramakrishnan B. She worked on real-time product inventions to be protected both in India and abroad. Within two years the company to its credit has bagged two National awards: the first one being "Indian Leaders for Industrial Empowerment" in the year 2016 and the second award being "Top Start-Up for IP Commercialization" in the year 2017. She was also been awarded "South Indian Women Achievers Award 'SIWAA'" in the year 2019 for her contribution to protecting IP Assets. She owns 12+ patents and she has been an IPR advisor and Indian Innovation Council (IIC) member for most of the well-reputed institutions in India. Though she took her place in entrepreneurship, her passion for teaching was still left unattended. Since 2022 November, she has been working as a temporary assistant professor at Mysore University School of Engineering affiliated with Mysore University in the Department of Artificial Intelligence and Data Science. She can be contacted at email: sgowthami12@gmail.com.

R. Venkata Siva Reddy 🆔 🔍 SC ◉ is a professor at REVA University with 32 years of teaching experience. Passionate IEEE volunteer having served IEEE Bangalore Section as Treasurer and Secretary. Received IEEE Bangalore section best volunteer award in the year 2013. Recipient of IEEE Education Society global award "IEEE Education Society Chapter Achievement Award in the year 2022". This year i.e., 2024, he is the General Chair of TALE-2024 conference sponsored by IEEE Education Society to be held in India from December 9th to 12th. Passionate about Amateur Radio activity and trustee of Indian Institute of HAM with a call sign VU3UFF. Seven students have been awarded PhD under my guidance as on date. Published more than 42 papers in peer reviewed international journals. He can be contacted at email: venkatasivareddy@reva.edu.in.

Mohammed Riyaz Ahmed ⓘ 🔗 SC ◖ (Senior Member, IEEE) received the B.E. and M.Tech. degrees in electronics and communication engineering and computer networking from Visvesvaraya Technological University, Belgaum, India, in 2007 and 2010, respectively, and the Ph.D. degree in electronics and communication engineering from JAIN University, Bengaluru, India, in 2016.,Since 2011, he has been with REVA University, Bengaluru, where he is currently an associate professor and an assistant director with the School of Multidisciplinary Studies. He is an advisor of IEEE EMBS Student Chapter and a mentor to IEEE ComSoc Student Chapter at REVA University. He is also the founder and the PI of the CPS Laboratory, REVA University. He is an advisor and the consultant to a number of higher educational institutes for their transformation into an Entrepreneurial University for knowledge economy. He is involved in investigation and teaching of design thinking as an enabler for individual, institutional, and international sustainable development. He has published more than 100 research papers in refereed journals and international conferences and as an invited speaker. His research interests include the computational cognitive neuroscience, WSNs, 5G and beyond, genomic data sequencing, tribology, neuromorphic engineering, spintronics, the IoT and edge computing, and green technologies. He has served as the TPC member for IEEE ICAECC Conference and a technical reviewer for seven IEEE journals and transactions along with numerous achieved journals and conference proceedings. He can be contacted at email: riyaz.pr@hkbk.edu.in.