International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 03 | March - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Dental Caries Detection System Using Machine Learning and IOT

Dr Brindha S[1], Mrs Priya D[2] , Mr Gokul Nath P[3] , Mr Gopi Krishna C[4] , Mr Jothiraj K[5],
Mohammed Dhanish H[6] ,Suriya Krishna G[7]

[1]Head of the Department, Computer Networking, PSG Polytechnic College, Coimbatore
[2]Lecturer, Computer Networking, PSG Polytechnic College, Coimbatore
[3,4,5,6,7] Students, Computer Networking, PSG Polytechnic College, Coimbatore

---***---

Abstract - Dental caries, a widespread oral health issue, necessitates early and accurate detection to ensure effective treatment and prevention. Traditional diagnostic methods often involve manual inspection and X-ray imaging, which can be time-consuming and subject to human error. This study introduces a cutting-edge approach by integrating machine learning with the YOLO (You Only Look Once) algorithm and leveraging the computational power of a Raspberry Pi to automate and enhance caries detection.

The research focuses on developing a real-time, portable system for detecting dental caries from intraoral images using a Raspberry Pi. The YOLO algorithm, known for its efficiency in real-time object detection, is employed to train a model capable of identifying and classifying carious lesions in dental radiographs with high precision. This study aims to advance dental diagnostic capabilities by combining machine learning and embedded systems technology. The use of the YOLO algorithm with Raspberry Pi not only enhances the accuracy and speed of caries detection but also makes this advanced technology accessible and practical for everyday clinical use.

Key Words: Dental caries, Open cv ,raspberry pi4 ,yoloV7

1.INTRODUCTION

Dental caries, commonly known as tooth decay, is one of the most prevalent oral health problems affecting people of all ages globally. Early detection is crucial for preventing the progression of caries, as it allows for timely treatment and minimizes the risk of severe dental complications. Traditionally, dental caries detection relies on manual examination by a dentist, or other dental imaging techniques. However, these methods can be subjective, time-consuming, and sometimes unable to detect early-stage caries. In recent years, the integration of Machine Learning (ML) and the Internet of Things (IoT) in healthcare has revolutionized the approach to disease detection, and dental care is no exception. A Dental Caries Detection System using these technologies offers an innovative, automated solution that enhances accuracy, speeds up the diagnostic process, and provides real-time monitoring and feedback. Machine Learning algorithms using (YOLO) can be trained to analyze dental images and detect signs of early-stage caries with higher precision than traditional methods. IoT-enabled devices, such as smart dental tools equipped with sensors, allow for the continuous collection of dental data, which can be analyzed in real-time and remotely via cloud platforms. This combination of ML and IoT can significantly improve the accuracy of diagnosis, reduce human error, and offer personalized, preventative care for patients.

In recent years, the integration of Machine Learning (ML) and the Internet of Things (IoT) in healthcare has revolutionized the approach to disease detection, and dental care is no exception. A Dental Caries Detection System using these technologies offers an innovative, automated solution that enhances accuracy, speeds up the diagnostic process, and provides real-time monitoring and feedback.

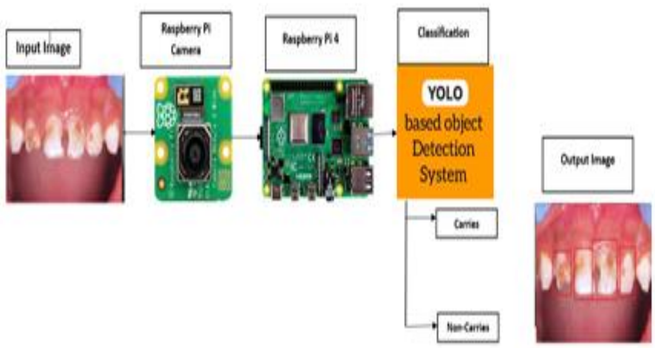


FIG 1 : BLOCK DIAGRAM

Machine Learning algorithms using (YOLO) can be trained to analyze dental images and detect signs of early-stage caries with higher precision than traditional methods. IoT-enabled devices, such as smart dental tools equipped with sensors, allow for the continuous collection of dental data, which can be analyzed in real-time and remotely via cloud platforms. This combination of ML and IoT can significantly improve the accuracy of diagnosis, reduce human error, and offer personalized, preventative care for patients.

This project highlights the implementation of OpenCV for real-time image processing on a Raspberry Pi. The system fetches live images from a camera and performs segmentation and analysis on them. OpenCV aids in image filtering, object recognition, and motion detection and tracking.

For surveillance, remote monitoring, and smart automations, this project serves as a economical solution using the processing power and compactness of Raspberry Pi. The object detection model served by YOLOv7 gets implemented. The entire system design along with the challenges and possible application is discussed. The results indicate that OpenCV integrated with Raspberry Pi leads to sophisticated image processing which encourages additional investigation in embedded systems.

1.1 MACHINE LEARNING

Machine Learning (ML) is a branch of artificial intelligence (AI) that enables computers to learn from data and improve their performance without being explicitly programmed. It involves using algorithms to recognize patterns and make decisions or predictions based on input data. There are three main types of ML: supervised learning, where the model is trained on labeled data; unsupervised learning, where the model discovers patterns in unlabeled data; and reinforcement learning, where the model learns through trial and error by receiving rewards or penalties for its actions. ML is widely applied in areas like recommendation systems, image recognition, and natural language processing.

2. PROPOSED MODEL

Machine learning workflow starts with problem understanding, meaning that there is a clear definition of the business or technical problem and identification of how machine learning can be used to address it. This phase determines the entire process foundation, with all its objectives aligned with business goals. Then there is data collection, where the applicable data is collected from diverse sources like databases, sensors, logs, or APIs. The quality and volume of this data are important to construct a solid model. After being collected, data annotation is done, particularly in supervised learning activities, where data is annotated to give the model appropriate output examples during training.

- Problem understanding – Define the business or technical problem and determine how machine learning can solve it.
- Data collection – Gather relevant data from various sources.
- Data annotation – Label or tag the data (if supervised learning is required).
- Data wrangling – Clean, transform, and preprocess the data to prepare it for modeling.
- Model development, training, and evaluation – Build the model, train it on data, and evaluate its performance using metrics.
- Model deployment and maintenance – Deploy the model to production and continuously monitor and update it as needed.

2.1 RASPBERRY PI4

Figure 2 : RASPBERRY P14

The Raspberry Pi is a series of credit card-sized computers developed in the UK by the Raspberry Pi Foundation with the intention of promoting the teaching of basic computer science in schools. The original Raspberry Pi and Raspberry Pi 2 were manufactured in several board configurations through licensed manufacturing agreements with Newark element14 (Premier Farnell), RS Components, and Egoman. These companies sell the Raspberry Pi online. Egoman produces a version for distribution solely in China and

Taiwan, which can be distinguished from others by their red coloring and lack of FCC/CE marks. The hardware remains the same across all manufacturers. The original Raspberry Pi is based on the Broadcom BCM2835 system-on-a-chip (SoC), which includes an ARM1176JZF-S 700 MHz processor, a VideoCore IV GPU, and was originally shipped with 256 MB of RAM, later upgraded (models B and B+) to 512 MB. The system features Secure Digital (SD) (models A and B) or MicroSD (models A+ and B+) sockets for boot media and persistent storage.

In 2014, the Raspberry Pi Foundation launched the Compute Module, which packages a BCM2835 SoC with 512 MB RAM and an MMC flash chip into a module for use in embedded systems. The Raspberry Pi 4 is a powerful single-board computer capable of running image processing algorithms and lightweight machine learning models. With its quad-core CPU and up to 8GB of RAM, it has sufficient processing power to handle moderate computational tasks, including dental image analysis for detecting caries and other oral health conditions.

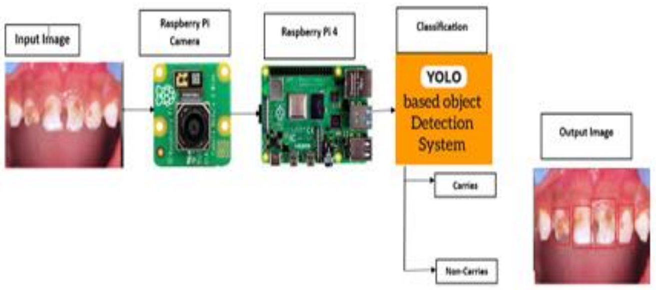


Figure 2.1 : DENTAL CARIES DETECTION SYSTEM

3. YOLO ALGORITHM

YOLO (You Only Look Once) real-time object detection algorithm, which is one of the most effective object detection algorithms that also encompasses many of the most innovative ideas coming out of the computer vision research community. Object detection is a critical capability of autonomous vehicle technology. It's an area of computer vision that's exploding and working so much better than just a few years ago. YOLO came on the computer vision scene with the seminal 2015 paper by Joseph Redmon et al. "You Only Look Once: Unified, Real-Time Object Detection," and immediately got a lot of attention by fellow computer vision researchers. Here is a TED talk by University of Washington researcher Redmon in 2017 highlighting the state of the art in computer vision.

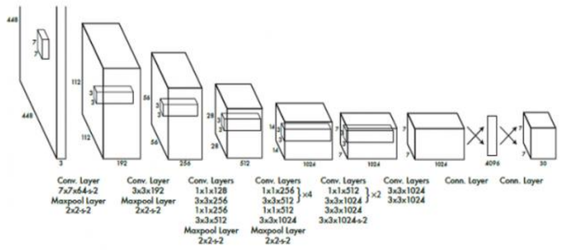


FIGURE 3 : YOLO ARCHITECTURE

Object detection is one of the classical problems in computer vision where you work to recognize *what* and *where* — specifically what objects are inside a given image and also where they are in the image. The problem of object detection is more complex than classification, which also can recognize objects but doesn't indicate where the object is located in the image. In addition, classification doesn't work on images containing more than one object. YOLO is popular because it achieves high accuracy while also being able to run in real-time. The algorithm "only looks once" at the image in the sense that it requires only one forward propagation pass through the neural network to make predictions. After non-max suppression (which makes sure the object detection algorithm only detects each object once), it then outputs recognized objects together with the bounding boxes. With YOLO, a single CNN simultaneously predicts multiple bounding boxes and class probabilities for those boxes. YOLO trains on full images and directly optimizes detection performance. This model has a number of benefits over other object detection methods:

- YOLO is extremely fast
- YOLO sees the entire image during training and test time so it implicitly encodes contextual information about classes as well as their appearance.
- YOLO learns generalizable representations of objects so that when trained on natural images and tested on artwork, the algorithm outperforms other top detection methods.

3.1 YOLO WORKING

The YOLO algorithm works by dividing the image into *N* grids, each having an equal dimensional region of SxS. Each of these *N* grids is responsible for the detection and localization of the object it contains. Correspondingly, these grids predict B bounding box coordinates relative to their cell coordinates, along with the object label and probability of the object being present in the cell. This process greatly lowers the computation as both detection and recognition are handled by cells from the image, but It brings forth a lot of duplicate predictions due to multiple cells predicting the same object with different bounding box predictions. YOLO makes use of Non Maximal Suppression to deal with this issue.

4. METHODOLOGY

The operating principle of a Dental Caries Detection System based on Machine Learning is the complete fusion of the latest data gathering, processing, and analyzing technologies. First, cameras like Raspberry Pi cameras capture high-definition images and oral health data concerning the patient's teeth in real time. The equipment continuously scans the oral cavity, collecting key information such as the state of the tooth surface and other oral health parameters.

The collected data is then fed into an ML system, where it undergoes preprocessing to refine its quality and relevance. Machine learning algorithms, particularly those based on YOLO (You Only Look Once), are applied to the data, extracting significant features and patterns to detect dental caries with high accuracy.

- Download the dataset from Mendeley Data

```
wget -O dental_caries_dataset.zip "https://data.mendeley.com/path/to/your/dataset"
unzip dental_caries_dataset.zip -d /home/pi/dental_caries_dataset
```

- Ensure the dataset follows YOLOv7 format

```
dental_caries_dataset/
|— images/
|   ├─ train/  # Training images
|   ├─ val/    # Validation images
|   ├─ test/   # (Optional) Test images
|— labels/
|   ├─ train/  # YOLO format labels for training
|   ├─ val/    # YOLO format labels for validation
|   ├─ test/   # (Optional) Test labels
```

Data collection in this research was crucial in establishing the groundwork for the overall machine learning process. Pertinent data was collected from credible sources as per the requirements of the problem. The procedure involved determining the appropriate type of data and sources such as databases, sensors, web APIs, or public datasets, depending on the problem type. Care was exercised to ensure that the data was complete, accurate, and representative of the domain being studied. At this phase, special attention was also given to the amount and diversity of data, ensuring that a comprehensive dataset was available for training effective models. Additionally, measures were taken to comply with data privacy laws and ethical practices. A preliminary analysis and exploration were performed to understand the data structure and identify potential issues such as missing values or inconsistencies that would need to be addressed in the next data preprocessing step.

4.1 DATA PROCESSING

Data preprocessing is an essential step in preparing the collected data for efficient analysis and model training. During this phase, the raw data is cleaned and transformed to maintain consistency, quality, and applicability to the machine learning process.Data preprocessing involves several key steps to ensure that the dataset is **clean, structured, and optimized** for machine learning algorithms.It starts with handling missing values using imputation or deletion methods, depending on the nature and amount of missing data. Duplicate records and unnecessary entries are identified and removed to enhance data integrity. Outliers are detected and appropriately handled to prevent them from skewing model performance.The data is then standardized or normalized to bring all features to a common scale, which is particularly important for algorithms sensitive to feature magnitude. Categorical variables are encoded using techniques such as one-hot encoding or label encoding, enabling the model to interpret them effectively. Additionally, feature selection and extraction techniques are applied to reduce dimensionality and retain only the most descriptive and relevant attributes. This ensures that the dataset is well-formatted, clean, and optimized for efficient and accurate model building.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 03 | March - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

4.2 MODEL SELECTION

On ce the data is prepared, the next step is to choose the appropriate machine learning algorithm or model. The choice of model depends on the type of problem (classification, regression, clustering, etc.), the size of the dataset, and the performance criteria set in the problem definition stage. For instance, if the goal is to classify emails as spam or not spam, a classification algorithm like logistic regression, decision trees, or support vector machines may be suitable. In contrast, if you are predicting a continuous variable, such as sales revenue, regression models like linear regression or random forests might be used. Sometimes, multiple models are tested, and the best-performing one is selected based on evaluation metrics. The selection of the right model is crucial because it directly influences the accuracy and effectiveness of the ML solution.

4.3 MODEL TRAINING

Once a model is selected, the next phase is training the model. In this stage, the model learns from the data by identifying patterns and relationships within the training dataset. The training process involves feeding the model input data and adjusting its internal parameters to minimize errors in predictions. Machine learning models use various optimization techniques, such as gradient descent, to fine-tune their weights. Depending on the complexity of the problem and the size of the data, training a model can be computationally intensive and time-consuming. It may also involve techniques like cross-validation to ensure that the model generalizes well to new, unseen data. The output of this stage is a trained model that can make predictions on new data based on the patterns it has learned.

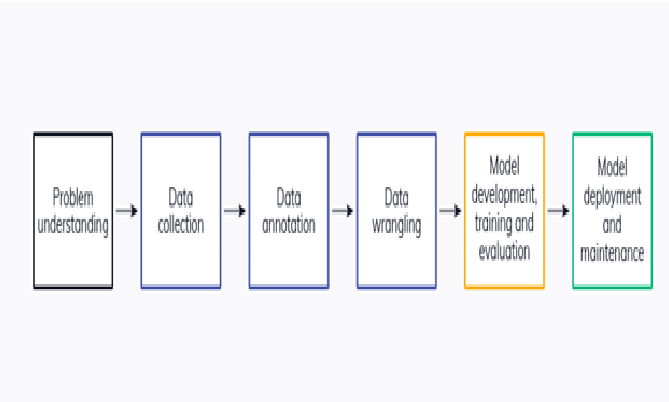


FIGURE 4.3 : WORKFLOW

4.4 MODEL EVALUTION

After the model is trained, it must be evaluated to ensure that it performs well on new, unseen data. This is where model evaluation comes into play. The performance of the model is assessed using a separate validation or test dataset that the model has not seen during training. Various metrics, such as accuracy, precision, recall, F1-score, and mean squared error, are used to measure the model's effectiveness depending on the problem type. For example, in a classification problem, accuracy may be the key metric, whereas in a regression problem, the focus may be on minimizing errors. In some cases, multiple evaluation metrics are used to provide a more holistic view of model performance. If the model performs well, it can be moved to the next stage. Otherwise, adjustments may be needed, including retraining the model, changing algorithms, or refining the features.

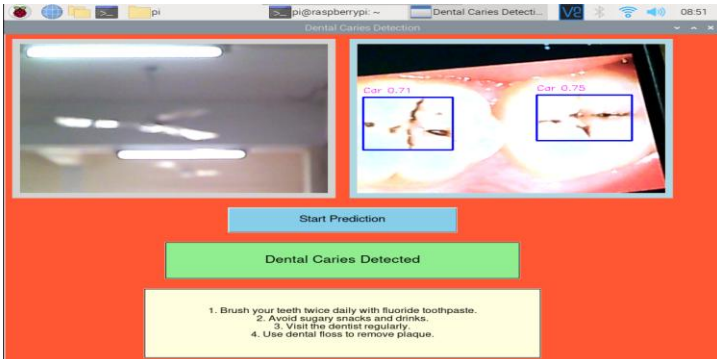
4.5 MODEL DEPLOYMENT

Once the model has been trained and evaluated, the next step is to deploy it into a production environment where it can make real-world predictions. Model deployment involves integrating the trained model with existing systems or applications to allow users to benefit from its predictions. This may require building an API, developing a user interface, or embedding the model within a larger software system. During deployment, care must be taken to ensure that the model can handle production-scale data and make predictions efficiently. Additionally, security and performance monitoring are important considerations during deployment to ensure that the model works as expected in a real-world setting.

4.6 MONITORING AND MAINTENANCE

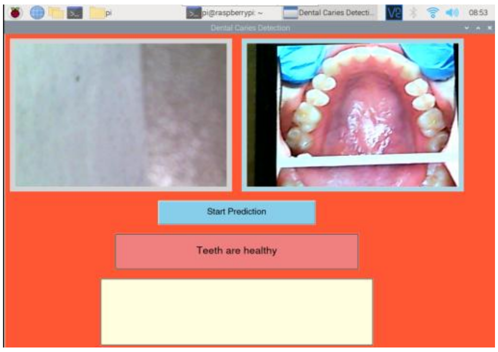
The final phase of the ML project life cycle is ongoing monitoring and maintenance, which is crucial for ensuring the long-term accuracy and reliability of machine learning models. Once deployed, these models must be continuously monitored to assess their performance as new data is introduced. Over time, models may experience "drift," where their predictions become less accurate due to changes in the underlying data or shifts in the environment. This drift can occur due to evolving user behavior, external factors, or changes in data distribution, making it necessary to regularly update, retrain, or even redesign the model to maintain its effectiveness. Continuous monitoring helps detect such issues early, allowing for timely interventions that prevent degradation in model performance. Automated monitoring tools can track key performance metrics, flag anomalies, and trigger alerts when accuracy declines. Additionally, feedback from users or stakeholders can provide valuable insights, leading to improvements, refinements, or adaptations of the model over time. By implementing a robust monitoring and maintenance strategy, organizations can ensure that their machine learning models remain accurate, relevant, and aligned with changing data patterns, ultimately enhancing their long-term effectiveness and reliability.

5. RESULT



Using **RealVNC Viewer** with a **Raspberry Pi 4** running an image detection system allows you to monitor and control the setup remotely from a PC, phone, or tablet. The Raspberry Pi connects to a camera and runs **YOLOv7** to detect objects in real time. With RealVNC Viewer, you can see a live feed where detected objects are marked with labels like "person" or "car." This setup is useful for security, surveillance, and industrial monitoring, especially in remote areas where physical access is limited. It runs at **5-10 FPS**, making it efficient for basic real-time detection. Additionally, it is a **low-cost and low-power** solution, ideal for home security, farms, and industrial sites.5. 1. RASPBERRY PI 4 OS DESKTOP

- Click on the Raspberry Pi Menu (top left corner).
- Go to Preferences > Raspberry Pi Configuration.
- Select the Interfaces tab.
- Enable VNC
- Click VNC to save the changes



5.1 CONNECT TO RASPBERRY PI WITH VNC VIEWER

- Open VNC Viewer on your client device.
- Enter the IP address of your Raspberry Pi in the VNC Viewer's address bar.
- Press Enter or click Connect. A prompt will ask for the username and password of your Raspberry Pi. Use:

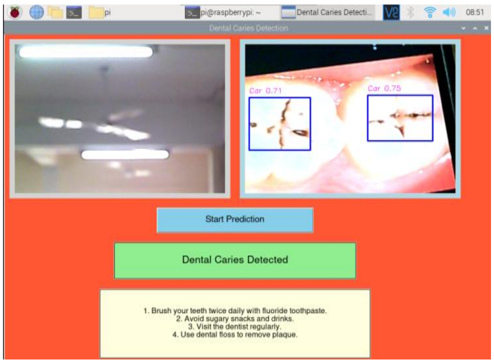
Once authenticated, you will be able to see and control the Raspberry Pi's desktop environment

FIGURE 5.2 : CONNECT TO RASPBERRY PI4 WITH REALVNC VIEWER

5. 2 USE CASES FOR REALVNC ON RASPBERRY PI4

- Remote Programming: Program or develop software on the Raspberry Pi without needing a physical monitor or keyboard.
- Server Management: Remotely manage a Raspberry Pi server for home automation, IOT, or as a web server.
- Teaching and Learning: Use Real VNC to help with remote learning, teaching students how to program or configure Raspberry Pi.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 03 | March - 2025 SJIF Rating: 8.586 ISSN: 2582-3930



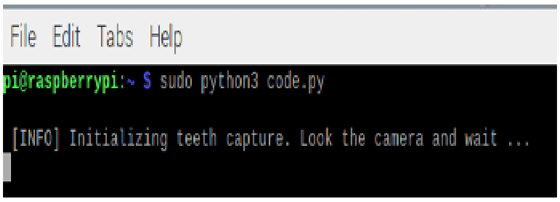


FIGURE 5.2: COMMAND USE FOR REALVNC VIEWER ON RASPBERRY PI4

6. CONCLUSION

The incorporation of Machine Learning (ML) and the Internet of Things (IoT) into a Dental Caries Detection System is a major breakthrough in dental care. The system offers a more precise, effective, and automated means of detecting dental caries, overcoming the disadvantages of conventional approaches like delayed diagnosis, subjectivity, and invasive procedures. With the use of IoT-enabled devices, including smart toothbrushes, intraoral cameras, and sensors, the system facilitates real-time monitoring and ongoing data collection, enabling early intervention and customized preventive care. The machine learning process—beginning with problem understanding, data gathering, annotation, preprocessing, model training, and deployment—is vital to the accuracy and reliability of the system. High-resolution images taken by Raspberry Pi cameras or other IoT sensors are analyzed to detect and classify dental caries, giving dentists useful information and minimizing manual errors in diagnosis. The use of machine learning algorithms improves the accuracy of diagnosis, providing consistent outcomes while minimizing human error.Although it has many benefits, there are challenges that need to be overcome to promote its use in clinical environments, including data privacy, model generalization, and hardware constraints. Ongoing enhancement of the quality of data, model training, and IoT integration will continue to perfect the system, turning it into a good asset for contemporary dentistry.

In summary, the Dental Caries Detection System is an innovative use of machine learning and IoT in the field of healthcare, leading the way to intelligent, quick, and accurate dental diagnosis, eventually enhancing patient care and overall oral health outcomes.

7. FUTURE WORK

Future work for the Dental Caries Detection System based on Machine Learning and IoT includes several improvements to enhance its efficiency, accuracy, and feasibility in real-world applications. One such area is improving deep learning models, where transformer-based architectures and enhanced CNN frameworks can be used to increase accuracy in identifying different stages of dental caries. Also, integrating multi-modal data sources like thermal imaging, 3D dental scans, and spectroscopy can provide a more comprehensive analysis of oral health conditions.

IoT integration and remote monitoring will also be a priority, facilitating improved wireless communication between dental imaging equipment, cloud storage, and mobile apps. Integration of 5G and blockchain technology can improve data transmission rates and security, providing real-time and tamper-proof patient records. The system can further be extended to tele-dentistry, where dentists can remotely diagnose and suggest treatments from real-time dental images.

• In addition, increasing and diversifying training datasets will be essential to enhance the model's robustness and generalizability. A greater dataset with images from diverse age groups, ethnicities, and dental conditions will reduce biases and enhance detection performance across diverse populations. Federated learning can also be pursued, enabling models to be trained on multiple healthcare institutions without exchanging sensitive patient data directly, hence ensuring privacy.By making early detection and prevention more accessible, affordable, and effective, these technologies will not only improve individual health outcomes but could also contribute to broader public health improvements worldwide. The continual advancements in AI, ML, IoT, and data analytics are likely to unlock even more opportunities for innovation in this space, leading to a more personalized, efficient, and connected approach to dental care.

ACKNOWLEDGEMENT

We sincerely express our gratitude to all those who contributed to the successful completion of this research on Dental caries detection system using machine learning and Iot. We would like to extend our heartfelt thanks to our mentors and advisors for their valuable guidance, insightful feedback, and continuous support throughout this study. Their expertise and encouragement have been instrumental in refining our understanding and improving the quality of this work. We are grateful to the open-source communities and developers who contributed to the advancement of machine learning and IoT technologies, making it possible for us to build an efficient and innovative system. Lastly, we extend our deepest appreciation to our family, friends, and colleagues for their unwavering support, motivation, and encouragement throughout this journey. Their belief in our work kept us motivated to strive for excellence. This project is a testament to the collaborative efforts of many individuals, and we sincerely appreciate everyone who contributed to its success.

REFERENCES

- 1. Young, D. A. et al. Te American Dental Association caries classifcation system for clinical practice: A report of the American Dental Association Council on Scientifc Afairs. J. Am. Dent. Assoc. 146, 79–86 (2015).
- 2. Lee, S. et al. Deep learning for early dental caries detection in bitewing radiographs. Sci. Rep. 11, 1–8 (2021).
- 3. Langlais, R. P., Skoczylas, L. J., Prihoda, T. J., Langland, O. E. & Schif, T. Interpretation of bitewing radiographs: Application of the kappa statistic to determine rater agreements. Oral Surg. Oral Med. Oral Pathol. 64, 751–756 (1987).
- 4. Rasib, M., Butt, M. A., Riaz, F., Sulaiman, A. & Akram, M. Pixel level segmentation based drivable road region detection and steering angle estimation method for autonomous driving on unstructured roads. IEEE Access 9, 167855–167867 (2021).
- K. Smith and D. Jones. “Automating Dental Caries Detection: A Machine Learning Approach.” Journal of Healthcare Informatics Research, Vol. 12, No. 1, 2020, pp. 34 46. This research covers the implementation of machine learning for dental diagnostics, particularly in automating the identification of caries through image analysis.

BIOGRAPHIES

Dr. S. Brindha is currently working as HoD, Computer Networking Department at PSG Polytechnic College, Coimbatore, TamilNadu. She joined PSG polytechnic College in the year 2000. Her research interests are in the area of Network Authentication and she has completed her doctorate in Information and Communication Engineering in the year 2015 from Anna University, Chennai. She has about 24 years of teaching and research experience. Performance Comparison of ASR ModelsShe has been coordinating the Autonomous Functioning activities for about 16 years. She has published many technical research papers and curriculum design related papers and won Best paper awards in Conferences. She has published many technical research papers and curriculum design related papers and won Best paper awards in Conferences. She has been instrumental in signing MoU with many companies and setting up industry oriented laboratories.

Ms D Priya is currently working as lecturer selection grade, Computer Networking Department at PSG Polytechnic College, Coimbatore, Tamil Nadu. She joined PSG polytechnic College in the year 2000. She has about 24 years of teaching and research experience in TCP IP Networking , wireless networks biometric security and machine learning. She has been instrumental in signing MoU with many companies and designing state of the art curriculum development. She has published many technical research papers and contributed chapters for technical publications.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 03 | March - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

GokulNath P (22DC15) is the Students of Diploma in Computer Networking, PSG Polytechnic College

Gopi Krishnan C (22DC16) is the Students of Diploma in Computer Networking, PSG Polytechnic College

Jothiraj K (22DC25) is the Students of Diploma in Computer Networking, PSG Polytechnic College

Mohammed Dhanish H (22DC27) is the Students of Diploma in Computer Networking, PSG Polytechnic College

Suriya Krishna G (23CH06) is the Students of Diploma in Computer Networking, PSG Polytechnic College