A Visionary Approach to Anemia Detection: Integrating Eye Condition Data and Machine Learning

Mrs. M. Asha Priyadarshini [1], Mrs. Sk. Salma [2], Damera Sailesh [3], Eda Manasa [4], G. Lakshmi Charan [5], Bussa Dinesh [6]

[1] Associate Professor, [2] Assistant Professor, [3,4,5,6] UG Scholar, damerasailesh@gmail.com
Department of CSE, Vignan's Lara Institute of Technology & Science,
Vadlamudi, Guntur, Andhra Pradesh.

Abstract. A low level of haemoglobin in the blood is known as anemia, and it can seriously harm important organs like the kidneys and heart. Conventional diagnostic techniques frequently require intrusive procedures, which causes anxiety in the patient and postpones care. This work uses images of the palpebral conjunctiva, which is known to show pallor in anemic people, to provide a unique, non- invasive method of detecting anemia. Our approach attempts to provide an easy-to-use way of early detection of at-risk individuals by conjunctival paleness analysis, allowing for prompt treatments. Using machine learning and integration of eye condition data, our innovative method guarantees quick, easy, and accessible anemia diagnosis for anyone. This advancement in technology has the potential to transform patient outcomes, promote global mental health, and change the way healthcare is delivered. We believe that with continued study and improvement, our method will have a major influence on the identification of anemia and open the door to improved health outcomes globally.

Keywords: Anemia, Haemoglobin, Palpebral conjunctiva, Non-invasive detection, Machine learning, Eye condition data, Early detection, Prompt treatment, Global health, Healthcare transformation

1 Introduction:

A common medical condition known as anemia is defined as a low amount of haemoglobin or red blood cells in the blood. These essential blood constituents are essential for carrying oxygen from the lungs to different tissues and organs. Depending on the underlying reason and degree of the anemia, several symptoms may occur. Fatigue, weakness, pale complexion, shortness of breath, cold hands, dizziness, headaches, and cold feet are typical indications and symptoms. Severe anemia can cause heart issues, cognitive decline, and immune system dysfunction, among other concerns. To avoid severe consequences and enhance the quality of life for patients, early detection and treatment are essential.

K. R. Madhavi et al. (eds.), *Proceedings of the International Conference on Computational Innovations and Emerging Trends (ICCIET 2024)*, Advances in Computer Science Research 112,
https://doi.org/10.2991/978-94-6463-471-6_75

Certain demographic groups—pregnant women, children, the elderly, and people with chronic illnesses—are disproportionately affected by anemia. Because they may lose blood during childbirth and have higher iron requirements throughout pregnancy, pregnant women are more susceptible to anemia. Anemia can also affect children, particularly in low- and middle-income nations where access to wholesome food and medical care may be restricted. Targeted interventions that enhance public health awareness, healthcare access, and nutrition are necessary to address anemia in these communities. Even though anemia is common and can have serious implications, it is frequently misdiagnosed or poorly managed, especially in areas with little resources and little access to medical care. Anemia affects more than a third of people worldwide, which highlights the importance of early discovery and appropriate treatment.

Conventional approaches to diagnose anemia usually entail intrusive procedures, including bone marrow biopsies and blood testing, which can be expensive, time-consuming, and uncomfortable for patients. Furthermore, it may be difficult to obtain certain diagnostic techniques, especially in areas with inadequate infrastructure and healthcare resources. As a result, non- invasive methods for diagnosing anemia that are easy to use, affordable, and simple are gaining popularity.

Technological developments have made it possible for creative methods of detecting anemia to emerge, leveraging the widespread use of cellphones and the strength of machine learning algorithms. With the advent of mobile health (mHealth) applications, people may now easily and remotely monitor their health condition, making them viable instruments for decentralized healthcare delivery. These apps take advantage of the cameras and sensors that come standard on smartphones to gather health-related data and give users immediate feedback. Particularly in settings with limited resources, mHealth platforms present the possibility of scalable and easily accessible screening options in the context of anemia identification. Through the utilization of mobile device's ease of use and connectivity, people can effortlessly take and share pictures of pertinent physiological indications, which can help with prompt diagnosis and treatment. Furthermore, to simplify patient administration and promote continuity of care, mHealth apps can interact with current healthcare systems. Leveraging mHealth technologies provides promise for increasing access to anemia screening and improving health outcomes globally as smartphone penetration rises.

To detect anemia, medical practitioners use a variety of diagnostic methods, such as serum iron indices, peripheral smears, and complete blood cell counts (CBC). Although these techniques offer insightful information on iron levels and blood composition, they frequently call for laboratory testing and blood samples, which may not be appropriate for many patients. Furthermore, some people may experience anxiety or anguish as a result of these procedures, which emphasizes the significance of non-invasive alternatives. Anemia-related data analysis has made extensive use of techniques including K- Nearest Neighbour (KNN), Naïve Bayes (NB), Random Forest (RF), and Support Vector Machine (SVM) in recent years. By using computational methods to categorize data points into several groups, these algorithms make it possible to identify anemia based on unique characteristics. The Machine Learning applications turn smartphone cameras into portable diagnostic instruments by allowing users to take pictures of their fingernails, conjunctiva, or other relevant body parts. These photos are then processed by sophisticated image processing algorithms to identify any indications of anemia.

While SVM has long been a popular option for conventional anemia detection techniques, more recent developments have led to the appearance of new algorithms with better performance and efficiency, like the XGBoost (XGB) classifier. The XGB classifier is an ensemble learning method that blends decision tree strengths to produce shorter computation times and increased accuracy. The XGB classifier can efficiently handle unbalanced datasets and derive insightful information from intricate data structures by utilizing gradient boosting techniques. With an emphasis on the XGB classifier, we present a novel method in this work for the non-invasive detection of anemia enhanced by machine learning techniques. Our study is to assess the efficacy of the XGB classifier in precisely diagnosing anemia based on conjunctival pictures, building upon earlier studies using SVM. We predict that the XGB classifier will perform better than the SVM in terms of accuracy, sensitivity, and specificity, providing a more stable and dependable anemia detection method.

In conclusion, anemia is a serious worldwide health issue with a wide range of demographic implications and possible side effects. Because of its widespread occurrence, there is an immediate need for easily accessible, non-invasive diagnostic methods to facilitate early detection and treatment. In resource-constrained environments, traditional procedures can provide difficulties, requiring the investigation of novel alternatives. The XGBoost classifier, in instance, is a machine learning technique that shows promise for improving the efficiency and accuracy of anemia detection. Using conjunctival pictures and sophisticated computational tools, our research aims to assess how well the XGB classifier outperforms more conventional approaches such as SVM. Our research is expected to aid in the creation of more dependable and scalable anemia detection instruments, which will ultimately enhance patient outcomes and lessen the prevalence of anemia worldwide.

2 Literature Survey:

In their work, Muljono, Sari Ayu Wulandari, et al. emphasise the conjunctival picture of the eye for non- invasive anemia detection with the use of the MobileNetV2 technique linked with the SVM algorithm [1]. Image preparation techniques such as data augmentation, imbalanced data handling, and data splitting were crucial when using the Eyes- defy-anemia dataset. The algorithms include SVM, MobileNetV2, and their combination, SVM-MobileNetV2 classification, which improves the effectiveness of diagnosis [1]. This integrated methodology, which provides a strong framework for early anemia identification through conjunctival imaging and has the potential to revolutionize diagnostic practices, highlights the need of utilizing machie learning techniques for medical image processing [1].

According to Endah Purwanti et al., anemia patients frequently have pallor in the palpebral conjunctiva, which is indicative of low haemoglobin levels [2] . This investigation was made easier by the dataset, which came from IEEE Dataport [2]. This work highlights the superiority of ResNet-50 in detecting anemia-related pallor in conjunctival pictures using AlexNet, ResNet-50, and MobileNetV2 [2]. These results highlight how deep learning models in medical image processing can be used to improve diagnostic capacities and possibly even improve patient outcomes by enabling early diagnosis and intervention in diseases like anemia [2].

In his discussion of the Random Forest (RF) and Decision Tree (DT) algorithms, Arvind Yadav et al. demonstrate how well these algorithms work in anemia analysis [3]. Information was gathered for this study from a variety of clinical databases, medical records, surveys, and publicly accessible datasets [3]. Decision trees, random forests, and SVM's were all included in the investigation; RF and DT were shown to be the most successful approaches. These results illustrate the value of using machine learning algorithms to diagnose anemia and demonstrate how RF and DT can enhance diagnostic precision and provide guidance to clinical decision-makers for better patient care and management [3].

Random Forest is the best algorithm for predicting anemia, according to Justice Williams Asare et al [4]. Obtaining 700 samples from various hospital sources was part of the data collection process. Important preprocessing stages included feature selection, balance, and normalization of the dataset. The importance of preparedness and data quality in improving the accuracy of predictive modelling is brought to light by this methodical process [4]. The promise of machine learning techniques in healthcare, especially in the prediction and management of illnesses like anemia, is highlighted by this study by utilizing Random Forest, which opens the door to more efficient diagnostic and treatment approaches [4].

According to Peter Appiahene et al., using smartphone cameras to identify anemia non-invasively is effective, with the CNN algorithm producing the highest accuracy [5]. This analysis was made easier by the dataset, which came from hospitals. In order to extract the region of interest, a modified circle technique was used to train an eye detection model on eye images. SVM, KNN, and CNN were among the algorithms used; CNN turned out to be the most accurate approach [5]. This study emphasizes how cutting-edge technologies, including deep learning algorithms and smartphone cameras, can be used for non-invasive medical diagnostics. This could completely change the way anemia is detected and screened for better healthcare results [5].

The eye conjunctiva, which beats out the palpable tongue, hand, and fingernails as the most trustworthy indicator for determining anemia, is discussed by Emmanuel Timmy Donkoh et al [6]. The dataset for study was gathered from ten health facilities located throughout Ghana [6]. The region of interest (ROI) was extracted and segmented using image preprocessing techniques. Three different algorithms were used: CNN, Decision Tree, and Naïve Bayes, with CNN working best [6]. These results highlight how important it is to use cutting-edge computational techniques, especially CNN, to detect anemia non-invasively and perhaps enhance diagnostic precision and patient outcomes [6].

Prakriti Dhakal et al. presented effective machine learning methods for the detection of anemia through the assessment of metrics and dataset features [7]. This assessment highlights the expediency of non-invasive techniques, such as machine learning algorithms, in tackling anemia diagnosis problems [7]. With an emphasis on anemia detection, a comprehensive review was carried out to outline the most recent developments and ideas in machine learning for healthcare. In this regard, a number of machine learning algorithms were examined and contrasted to shed light on their effectiveness and their uses in boosting diagnostic capacities for better patient outcomes [7].

Bheem Sen et al. emphasize the evaluation of machine learning classifiers for classification [8]. These classifiers include random forest, logistic regression, naïve Bayes, and support vector machine (SVM) [8]. Metrics for both accuracy and computing efficiency were used to analyze their performance. This evaluation emphasizes how crucial it is to choose the right classifiers for a given task while weighing computing requirements and accuracy [8]. Furthermore, preliminary preprocessing of microscopic pictures was carried out, emphasizing the importance of data preparation in improving machine learning algorithms' performance [8]. These results aid in the improvement of preprocessing techniques and classifier selection for effective and precise classification across a range of domains [8].

In their research, Manish Jaiswal et al. demonstrate supervised machine learning algorithms to predict anemia using Complete Blood Count (CBC) data from pathology centers [9]. Interestingly, the accuracy of the Naive Bayes technique was higher than that of the Random Forest and Decision Tree algorithms, demonstrating its superiority in the prediction of anemia [9]. The efficacy of Naive

Bayes was further confirmed by comparing performance indicators. The process of controlling outliers using data cleaning techniques was essential in preparing the dataset for precise modelling [9]. These results highlight how important it is to choose algorithms carefully and use reliable data preprocessing techniques when optimizing predictive models for medical diagnostics, especially when using CBC data to predict anemia in pathological settings [9].

When Pooja Pradeep Dalvi et al. discuss the ensemble methods used, the stacking ensemble approach beats out bagging, voting, adaboosting, and bayesian boosting [10]. On the other hand, of all the classifiers, K-Nearest Neighbour has the lowest accuracy. Together, these techniques improve accuracy by utilizing predictions from several base classifiers [10]. By combining various classifier outputs, stacking, bagging, voting, adaboosting, and bayesian boosting enhance prediction performance [10]. The effectiveness of ensemble approaches in maximizing model accuracy is highlighted by this analysis, which provides a solid method for utilizing machine learning in predictive analytics and decision-making procedures [10].

3 Proposed Methodology:

Dataset Collection:

A dataset of 3000 photos for each class— anemia and non-anemia—was acquired from Kaggle in order to compile the dataset. The RGB format photos in the collection have 64 by 64 pixel size and three color channels (64, 64, 3). Because all of the photographs are saved in the JPG format, typical image processing programmes and libraries will work with them. To enable thorough model training and evaluation, a balanced representation of both anemic and non-anemic patients is maintained in the selection of 3000 images for each class. The pictures show a range of visual features, including normal eye appearances and various anemia- related pallor presentations in the palpebral conjunctiva.

The RGB format makes it possible for color information to be represented in the photos, giving anemia detection algorithms more context. The photograph's 64 × 64 pixel size strikes a compromise between computing effectiveness and the retention of pertinent visual information, guaranteeing reasonable processing speeds while maintaining enough resolution for precise analysis. The dataset is homogenous due to its use of a standardized image format (JPG) and constant dimensions throughout all samples, making it easy to integrate into machine learning workflows. As a result of this uniformity, preprocessing operations like data augmentation, image resizing, and normalization become more efficient and productive, which benefits the subsequent model training processes. Overall, the dataset collection process ensures the availability of a comprehensive and wellstructured dataset tailored for training and evaluating machine learning models for non-invasive anemia detection using conjunctival images.

Data Preprocessing:

Data preprocessing is a critical step in preparing the dataset for machine learning model training, ensuring optimal performance and accuracy. In data preprocessing, the images are first loaded and resized to a standard size of 64 x 64 pixels using OpenCV's `cv.resize()` function. This resizing ensures uniformity in the dimensions of all images, reducing computational complexity and facilitating consistent feature extraction across the dataset. Each image is then associated with a numerical label representing its class (0 for anemia, 1 for non-anemia) based on the directory structure. The dataset is subsequently shuffled to introduce randomness in the arrangement of samples, minimizing biases during model training. The features (images) and labels are extracted from the dataset and stored in separate lists, `x_data` and `y_data`, respectively. These lists are then converted into NumPy arrays using `np.array()` for efficient computation and compatibility with machine learning frameworks. Finally, pixel values in the images are normalized to the range [0, 1] by dividing each pixel value by 255. Normalization standardizes the input data, facilitating faster convergence during model training and improving overall performance.

Implementation:

The methodology centers on employing the powerful XGBoost model for precise anemia detection. In this process, we initialize and train an XGBClassifier on our training dataset, laying the groundwork for subsequent iterations. Harnessing the prowess of gradient boosting algorithms, our XGBoost model adeptly discerns complex patterns within the data, thereby elevating its predictive accuracy to new heights. Once trained, the model seamlessly extends its predictive prowess to generate insightful predictions for our testing dataset, showcasing its remarkable ability to generalize effectively to previously unseen instances. By meticulously saving the trained model, we not only ensure its accessibility but also open avenues for future deployments, thereby fostering seamless integration into clinical practice. Through a comprehensive evaluation, featuring metrics such as accuracy score, classification report, and confusion matrix, we gain a profound understanding of our model's predictive acumen, precisely quantifying its efficacy in distinguishing between anemia and non- anemia cases. Furthermore, an illustrative example prediction provides a vivid demonstration of the model's practical utility, offering invaluable insights into its decision-making process within real-world clinical scenarios.

By capitalizing on the inherent strengths of the XGBoost model, our methodology presents a formidable and precise solution for anemia detection. Each meticulously orchestrated step within the model serves as a critical cog in the wheel, effectively paving the path for accurate prediction of anemia based on input images. This approach not only streamlines diagnostic procedures but also stands as a beacon of hope for improved healthcare outcomes, ensuring timely interventions and treatments for those in need.

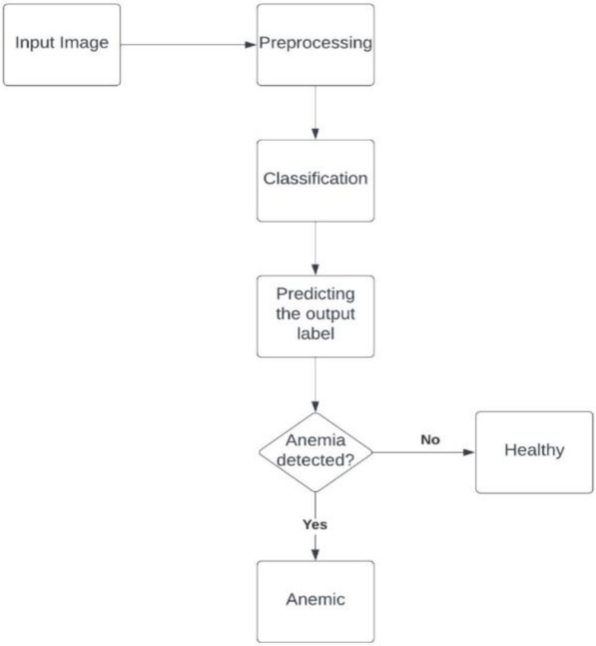

Fig. 1. Working Architecture of the model

S. No	Evaluation Metrics	XGB Model
1.	Accuracy	93%
2.	Precision	93%
3.	Recall	96%
4.	F1 Score	94%

Table 1. Evaluation Metrics of Model

4 Results:

This report's evaluation metrics offer a thorough examination of the model's functionality. Recall, which measures the percentage of properly predicted true positives out of all actual positives, and precision, which shows the percentage of true positive predictions out of all positive predictions for each class, provide information about the accuracy and efficacy of the model. As the harmonic mean of recall and precision, the F1-score guarantees a fair evaluation of performance. Furthermore, the number of samples in each class is indicated in the support column, which further contextualizes the assessment metrics. The model's overall good precision, recall, and F1-score validate its ability to correctly categorize cases as either anemia or non- anemia. Moreover, comprehensive examination employing cutting-edge computer methods ascertains whether submitted photos exhibit anemia symptoms or traits linked to optimal health. Further clarifying the distribution of true and predicted class labels, the heatmap visualizes the confusion matrix and offers important information into the assessment of the XGB Classifier model.

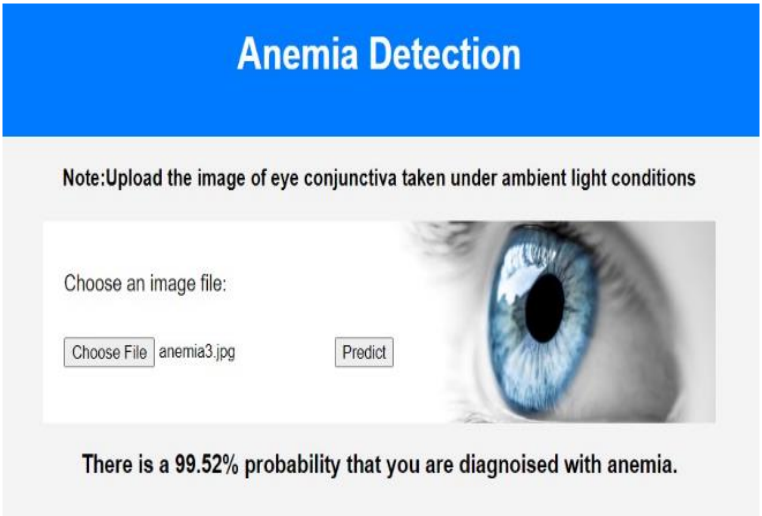

Fig. 2. Model predicting uploaded image as anemic

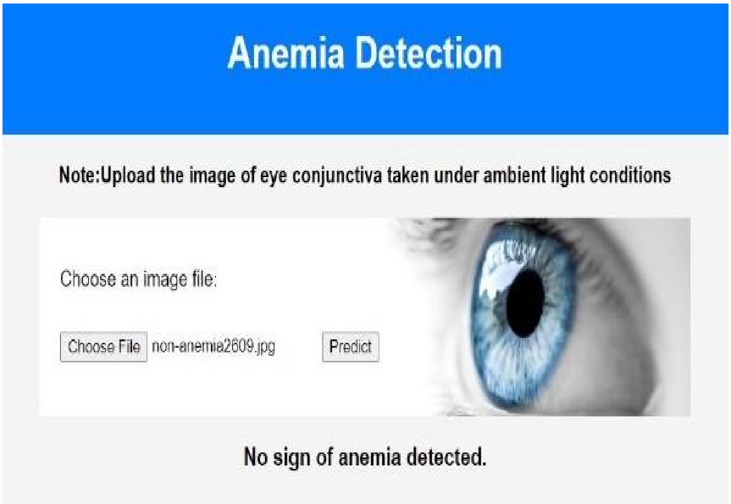

Fig. 3. Model predicting uploaded image as healthy

Fig. 4. Confusion matrix for XGB Classifier model

5 Conclusion:

In conclusion, this research endeavors to revolutionize the landscape of anemia detection through the integration of advanced machine learning techniques, specifically leveraging the XGBoost model. By meticulously analyzing conjunctival images, our methodology presents a non- invasive and highly accurate approach to identifying individuals at risk of anemia. Through the meticulous evaluation of our model's performance, including precision, recall, and F1-score metrics, we have demonstrated its remarkable efficacy in accurately classifying anemia and non- anemia cases. The practical utility of our model is further underscored by an illustrative example prediction, shedding light on its decision-making process within clinical settings. This research not only provides a robust framework for anemia detection but also paves the way for improved healthcare outcomes by enabling timely interventions and treatments. Looking ahead, continued advancements in machine learning algorithms and image analysis techniques hold immense promise for further enhancing the accuracy and accessibility of anemia detection methods, ultimately leading to better health outcomes for individuals worldwide.

6 Future Scope:

In addition to the current findings, there exists a wealth of potential avenues for future exploration in the realm of anemia detection. Further research could delve into the integration of multi-modal data sources, such as clinical parameters and genetic markers, to enhance the predictive capabilities of our models. Additionally, the development of mobile applications leveraging our non-invasive approach could facilitate widespread adoption and real-time monitoring of anemia status. Exploring the application of transfer learning techniques could enable the adaptation of our models to diverse populations and healthcare settings, fostering greater inclusivity and accessibility. Moreover, collaborative efforts between medical professionals and data scientists could lead to the creation of comprehensive decision support systems, empowering clinicians with actionable insights for personalized patient care. Finally, continued advancements in machine learning algorithms and computational techniques hold immense promise for refining and optimizing anemia detection methods, ultimately driving improvements in public health outcomes on a global scale.

7 References:

1. Jaiswal, Manish, Anima Srivastava, and Tanveer J. Siddiqui. "Machine learning algorithms for anemia disease prediction." Recent Trends in Communication, Computing, and Electronics: Select Proceedings of IC3E 2018. Springer Singapore, 2019.
2. Asare, Justice Williams, et al. "Iron deficiency anemia detection using machine learning models: A comparative study of fingernails, palm and conjunctiva of the eye images." Engineering Reports (2023): e12667.
3. Dalvi, Pooja Tukaram, and Nagaraj Vernekar. "Anemia detection using ensemble learning techniques and statistical models." 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT). IEEE, 2016.
4. Sen, Bheem, et al. "Machine learning based Diagnosis and classification of Sickle Cell Anemia in Human RBC." 2021 Third International Conference on Intelligent

Communication Technologies and Virtual Mobile Networks (ICICV). IEEE, 2021.

5. Asare, Justice Williams, Peter Appiahene, and Emmanuel Timmy Donkoh. "Detection of anemia using medical images: A comparative study of machine learning algorithms–A systematic literature review." Informatics in Medicine Unlocked (2023): 101283.
6. Dhakal, Prakriti, Santosh Khanal, and Rabindra Bista. "Prediction of anemia using machine learning algorithms." AIRCC's International Journal of Computer Science and Information Technology (2023): 15-30.
7. Bauskar, Shubham, Prakhar Jain, and Manasi Gyanchandani. "A noninvasive computerized technique to detect anemia using images of eye conjunctiva." Pattern Recognition and Image Analysis 29 (2019): 438-446.
8. Çil, Betül, Hakan Ayyıldız, and Taner Tuncer. "Discrimination of β-thalassemia and iron deficiency anemia through extreme learning machine and regularized extreme learning machine based decision support system." Medical hypotheses 138 (2020): 109611.
9. Roychowdhury, Sohini, et al. "Computer aided detection of anemia-like pallor." 2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI). IEEE, 2017.
10. Asare, Justice Williams, et al. "Application of machine learning approach for iron deficiency anemia detection in children using conjunctiva images. "Informatics in Medicine Unlocked (2024):101451.
11. Annapurna, B., Asha Priyadarshini Manda, A. Clement Raj, R. Indira, Pratima Kumari Srivastava, and V. Nagalakshmi. "Max 30100/30102 sensor implementation to viral infection detection based on Spo2 and heartbeat pattern." Annals of the Romanian Society for Cell Biology (2021): 2053-2061.
12. Sendak, Mark P., et al. "Real-world integration of a sepsis deep learning technology into routine clinical care: implementation study." JMIR medical informatics 8.7 (2020): e15182.
13. Chitteti, Chengamma, and K. Reddy Madhavi. "Taylor African vulture optimization algorithm with hybrid deep convolution neural network for image captioning system." Multimedia Tools and Applications (2024): 1-19.
14. Sangle, Sanchita, et al. "Conference report: Dhara-Vision Ayurveda 2047." Journal of Ayurvedaand Integrative Medicine 15.1 (2024): 100848.
15. Raman, Rohan Kumar, et al. "Reconnoitering Precision Agriculture and Resource Management: A Comprehensive Review from an Extension Standpoint on Artificial Intelligence and Machine Learning."
16. Baghel, Ruchi, et al. "Integration of epigenomics and metabolomics: from biomarkers discovery to personalized medicine." Epigenetics and metabolomics. Academic Press, 2021. 31-7
17. Pahuja, Rahul, et al. "A Dynamic approach of eye disease classification using deep learning and machine learning model." Proceedings of Data Analytics and Management: ICDAM 2021, Volume 1. Springer Singapore, 2022.
18. Raju, S. Viswanadha, A. Vinaya Babu, G. V. S. Raju, and K. R. Madhavi. "W-Period Technique for Parallel String Matching." IJCSNS 7, no. 9 (2007): 162.
19. K.Venkateswara Rao, ”Disease Prediction and Diagnosis Implementing Fuzzy Neural Classifier based on IoT and Cloud”, International Journal of Advanced Science and Technology (IJAST), ISSN : 2005-4238, Vol-29 Issue-5, May 2020,

Page No: 737-745.

20. K.Venkateswara Rao, "Research of Feature Selection Methods to Predict Breast Cancer", International Journal of Recent Technology and Engineering(IJRTE), ISSN : 2277-3878, Vol-8 Issue-2s11, Sep 2019, Page No: 2353-2355.
21. Desanamukula, Venkata Subbaiah, M. Asha Priyadarshini, D. Srilatha, K. Venkateswara Rao, RVS Lakshmi Kumari, and Kolla Vivek. "A Comprehensive Analysis of Machine Learning and Deep Learning Approaches towards IoT Security." In 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 1165-1168. IEEE, 2023.