Smart IoT-Enabled pH Strip Reader for Simple and AccurateAnalysis of Cervical Mucous pH

Nitya Rajan[1*], Muhila S, Pranay Parikh, and Anish Raj Arisetty

Vellore Institute of Technology, Vellore, Tamil Nadu

nityarajan02@gmail.com

Abstract. The pH analysis of cervical mucous plays a significant role in providing valuable health insights for women. Accurate measurement of this pH metric is crucial for effective tracking and monitoring of reproductive health. However, current methodologies often present challenges in terms of confusion in test strip readings and susceptibility to external factors suchas light fluctuations. In this study, we propose a robust and simplified approach for measuringcervical mucous pH, leveraging the integration of Internet of Things (IoT) technology and TinyML. Our setup ensures accurate readings, minimizing ambiguity, and offering a reliable means for women to effortlessly monitor their cervical mucous pH. We conducted experimentsto compare the accuracy of our device with human interpretation of universal pH strips. The results show our setup demonstrates superior accuracy compared to laymen's interpretationsof universal pH strips with small increments in pH (0.5 pH difference). This methodology empowers women to gain deeper insights into their reproductive health, aiding in informed decision-making and improving overall well-being.

Keywords: pH, Cervical Mucous, TinyML, Smart IoT, Random Forest Regression

1 Introduction

There have been a few attempts at assessing pH via image analysis. Loh et Al. [6] proposed a pH reader that is automated by a mobile phone in 2011, by using Edge detection to isolate the pH strip and further performing K-means clustering for classification. Kim et Al. [4] improved on this concept by constructing a light box around the pH strip in order to render the system less vulnerable to ambient light fluctuations, improving accuracy. Dello et Al. applied the methodology proposed by Kim et Al for pH analysis of Beer samples with promising results in 2020 [1]. Da Silva et Al. applied the method to water quality analysis in the same year[12]. In 2022, Elsenety et Al. [2] proposed a simple and accurate implementation of a mobile phone-based strip reader using a large library of pH values with a 0.1 increment. In the same year, Pazzi et Al. approached pH

C. Kiran Mai et al. (eds.), *Proceedings of the Fourth International Conference on Advances in Computer Engineering and Communication Systems (ICACECS 2023)*, Atlantis Highlights in Computer Sciences 18,
https://doi.org/10.2991/978-94-6463-314-6_22

strip reading through RGB index analysis of paper-based analytical devices via an RBG sensor [9]. Advancements in machine learning algorithms, particularly regression techniques, provide a promising approach to address the accuracy issues associated with pH strips. By effectively mapping the relationships between the RGB values obtained from the pH strips and their corresponding pH values, machine learning models can enhance the accuracy of pH measurement. Notably, these algorithms can be optimized to run efficiently on small commercial microcontrollers, such as the Nodemcu or Arduino, using TinyML libraries, enabling the creation of a standalone device.

2 Importance of pH measurement in cervical mucous

Cervical mucous, also known as cervical mucus, is a fluid secreted by the cells of the cervix within the female reproductive system. It plays a vital role in the reproductive process by facilitating sperm transport, providing lubrication during sexual activity, and acting as a protective barrier against pathogens.

The measurement of the pH of cervical mucous holds significant value and insights for various reasons. One of the prominent applications is its role as an indicator for Bacterial Vaginitis (BV), a common vaginal infection. Normally, the vagina maintains a moderately acidic pH level ranging from 3.5 to 4.5[7]. However, in cases of BV, the pH of cervical mucous becomes more alkaline, serving as a diagnostic marker for the condition [5]. Often, BV is falsely treated using OTC antifungal medication. Appropriate diagnosis can lead to correct treatments and better patient outcomes [10].

Furthermore, the pH of cervical mucous has been linked to unexplained fertility. Sperm survivability within the vaginal and cervical canal is influenced by the pH value, as an acidic environment can be detrimental to sperm[8]. Therefore, accurately analyzing the pH of cervical mucous becomes crucial for tracking fertility and understanding factors affecting sperm viability.

The regulation of pH in the vaginal environment is controlled by the natural microbiota of the vagina and hormonal factors [3]. Consequently, analyzing the pH of cervical mucous serves as a usefulindicator for various aspects of vaginal health.

Currently, interpreting vaginal pH strips can be challenging, potentially leading to inaccuracies in pH measurements. Additionally, a more precise analysis of vaginal pH and its variations throughoutthe menstrual cycle can provide further insights into vaginal health and menstrual cycle patterns. In light of these considerations, we propose the development of a portable and user-friendly Internet of Things (IoT) enabled vaginal pH strip reader. This innovative solution aims to simplify and enhance the accurate measurement and tracking of vaginal pH, empowering women to monitortheir vaginal health effectively. The integration of IoT technology further enables seamless data tracking and analysis, facilitating comprehensive insights into pH variations and their implications.

3 Proposed Methodology

The proposed standalone device encompasses a closed box design that minimizes errors in RGB readings by providing a controlled environment. By mitigating the impact of fluctuating lighting conditions and ensuring accurate strip centering, the device delivers more reliable and precise pH measurements. The basic flow and structure of our proposed setup can be seen in Figure 1.

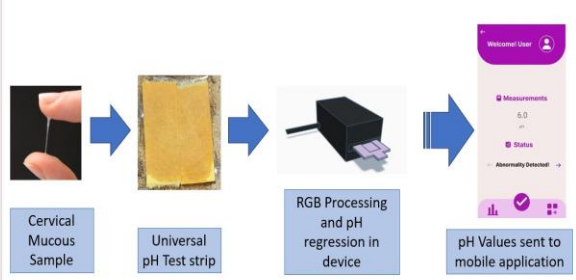


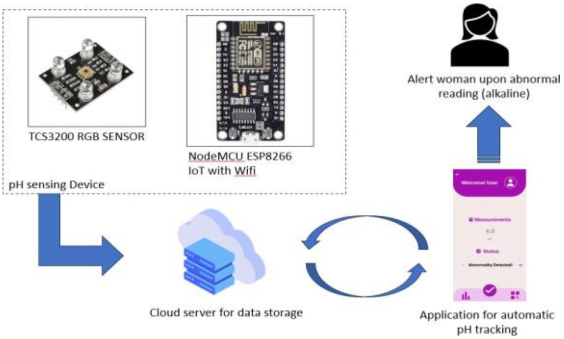
Fig. 1. Structure of the proposed setup

Additionally, leveraging the Internet of Things (IoT) technology, the device can be integrated with a user-friendly app interface, offering seamless interaction, data storage, and analysis options. This IoT implementation enables users to access pH measurements remotely, track vaginal pH trends, and make informed decisions based on the stored data (Fig 2).

The setup employed for pH sensing is designed with the goal of minimizing potential sourcesof error, such as variations in background light and inconsistent distances between the sensor and the pH strip. The setup ensures a dark environment and a constant distance, which contributes to stabilizing and enhancing the accuracy of the readings obtained.

The TCS3200 sensor, developed by Texas Advanced Optoelectric Solutions, is a versatile programmable color light-to-frequency converter utilized in the pH sensing device for RBG value reading. This integrated circuit combines configurable silicon photodiodes and a current-to-frequency converter on a single monolithic CMOS chip. It provides an output in the form of a square wave with a 50 percent duty cycle, where the frequency is directly proportional to the intensity of light it detects.

Fig. 2. Synoptic of pH tracking device for cervical mucous tracking



A TCS3200 sensor is affixed to the top of the box or enclosure. This placement ensures that the sensor is positioned correctly to capture the RGB values of the pH strip during the measurement process. Figure 2 represents a general synoptic of our setup and data flow mechanism. To develop a reliable machine learning model for pH sensing, it is essential to establish a comprehensive database that correlates the RGB values obtained from the TCS3200 sensor with the corresponding pH readings. By directly collecting readings using the devised setup, the need for additional calibration, such as relying on a given reference sheet, is eliminated. This is crucial as the pH paper's properties, such as moisture content, can significantly impact the readings and introduce inaccuracies.

To construct the database, a systematic approach was employed (see Fig 3). Multiple phosphate and acetate buffers were prepared, covering a pH range from 3.5 to 7.0 with an increment of 0.5pH units. This reflects the range of values that cervical mucous can be measured as. These buffers served as the reference standards for the subsequent data collection process. A cotton swab was used to apply a small sample of each buffer onto the pH paper. It was ensured that sufficient time was allowed for the paper to dry, as variations may arise during the drying process, which can affect the measured values.

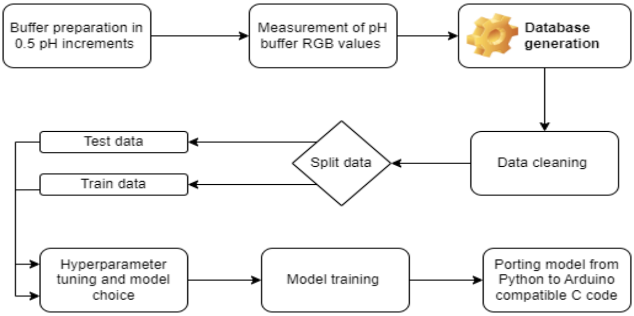


Fig. 3. A flowchart of the employed data collection methodology

The data collection procedure involved taking readings from the TCS3200 sensor for a durationof 1.5 minutes. This duration was chosen to account for any fluctuations or changes in the pH paper's response as it dried. By observing the pH value for this extended period, it aimed to capture a more accurate representation of the pH within the desired timeframe of 1.5 minutes.

Once the data collection phase was complete, the obtained dataset underwent a thorough cleaning and processing procedure (Fig.3). This involved removing any outliers or inconsistencies that could introduce noise or bias into the dataset. Additionally, the data was formatted and transformedas necessary to ensure compatibility with the machine learning model being employed. Figure 4. illustrates the distribution of measured RGB values for each pH. The upward streak indicates thedrying of the paper, resulting in higher RGB values.

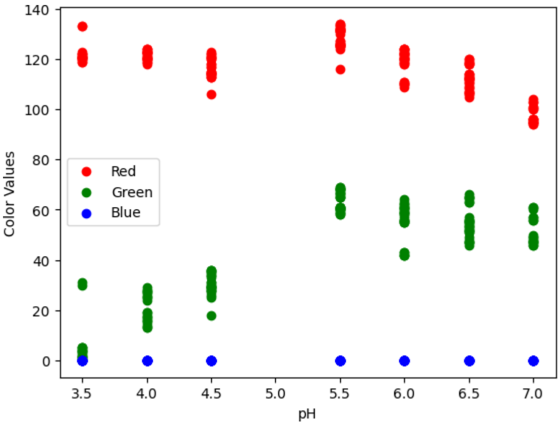


Fig. 4. Scatterplot of RGB values measured for each pH

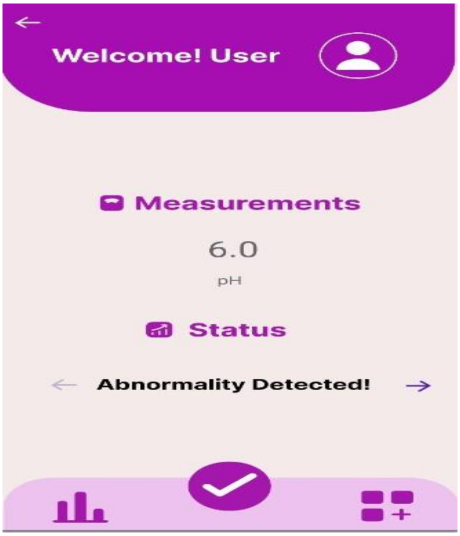
After dividing the collected dataset into training and testing sets, we employed gridSearchCV, a technique for hyperparameter tuning, to identify the optimal parameter values for four different regression models: Decision Tree Regressor, Random Forest Regressor, Linear Regressor, and XGB Regressor. The evaluation criterion used to compare the models was the Root Mean Squared Error (RMSE).

Figure 5a illustrates a CAD model of the setup, providing a visual representation of its components,and 5b shows our implementation using cardboard. The setup involves a simple board, depicted in purple, which features an indentation specifically designed to hold the pH strip securely during readings. This ensures that the strip remains stationary and avoids any unintended movement that could impact the accuracy of the measurements.

(a) CAD Model (b) Cardboard implementation

Fig. 5. Setup design and implementation

To facilitate ease of use and streamline the process, the board with the pH strip can be conveniently inserted into a slit in a box or enclosure. This allows for efficient and straightforward readings tobe taken without any additional complexity. Finally, a wifi based connection was established using Firebase. The measured pH values were transmitted to a linked app, where the values were tracked and presented to the user, as seen in the wireframe in Figure 6.



Welcome! User

Measurements

6.0

pH

Status

Abnormality Detected!

Fig. 6. Wireframe of App interface with Alert mechanism

The user was flagged if the measured pH exceeded the normal range of vaginal pH, that is 3.5-5pH.

4 Experimental Results

Table 1. RMSE Values for Different Models

Model	RMSE
Decision Tree Regressor	0.4951
Random Forest Regressor	0.2372
Linear Regression	0.9325
XGB Regressor	0.3966

Based on the results obtained as seen in Table 1, the Random Forest Regressor exhibited the lowest RMSE among the four models, indicating superior predictive performance. Therefore, we selected the Random Forest Regressor as the most suitable model for our pH sensing application. The feature importance graph (Fig 7) shows that the model is most sensitive to variation in green values. In the next step, we ported the Random Forest Regression model from Python to Arduino-compatible C code using the micromlgen library from Eloquent Arduino. The Random Forest Regression implementation in Arduino required zero bytes of RAM since it hard-codes allthe constituent decision trees. However, it does utilize as many bytes as the number of classes for storing the votes, which is negligible in our case [11].

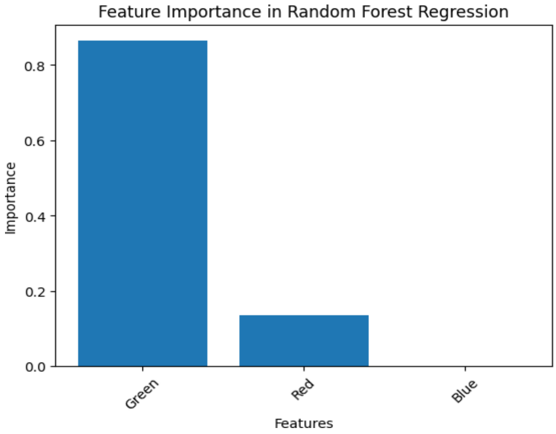


Fig. 7. Feature Importance Graph of Random Forest Regression

Table 2. Performance Metrics of Random Forest Regression

Metric	Value
RMSE	0.1595
Accuracy	0.97983

This allows us to run the regression of the NodeMCU. The Random Forest regression model achieved an RMSE (Root Mean Squared Error) of 0.1595 and an accuracy of 0.97983 was achieved for the model as shown in Table 2. These performance metrics, the low RMSE and high accuracy, suggest that the Random Forest regression model successfully captures the underlying patterns and relationships between the RGB values from the TCS3200 sensor and the corresponding pH readings. This indicates that the model is capable of accurately estimating the pH levels based on the provided RGB inputs, making it a reliable tool for pH-sensing applications.

In order to analyse whether our device improved the accuracy of reading test strips we asked 3 laymen test subjects to interpret various pH strips from pH 3.5 to 7.0 with increments of 0.5 pH and provided them with the reference chart. We observed a large deviation between the subjects, showing confusion in accurately reading universal pH strips with small differences in pH (Table 3).

Table 3. True and Predicted pH Values for Subjects A, B, and C

True pH	Predicted pH	Subject A	Subject B	Subject C
3.5	3.6	4	3.5	4
4	4	3.5	4.5	3.5
4.5	4.3	4.5	4	4.5
5	5.2	5.5	5	5
5.5	5.5	5	6	6
6	6.2	6	5.5	5.5
6.5	6.4	7	6.5	6.5
7	7	6.5	7	7

The subjects have an average error of ±0.3pH compared to the true pH. In contrast, our device demonstrated superior accuracy with an average ±0.1pH error compared to the true pH.

5 Conclusion

The analysis of the pH of cervical mucous holds substantial potential in offering valuable health insights to enhance the well-being of women. The availability of a simple and precise method to measure this pH, minimizing confusion in test strip readings, would greatly facilitate the monitoring of this essential metric. Our proposed setup, designed to provide accurate readings and mitigate the influence of external factors such as light fluctuations, establishes a robust framework for obtaining reliable results. It achieves greater accuracy than laymen's readings. By integrating the Internet of Things (IoT) technology, we envision that this methodology can be effectively employed to enable women to effortlessly track their cervical mucous pH levels.

Such an approach would empower women to gain deeper insights into their reproductive health,facilitating informed decision-making and contributing to an improved quality of life

References

1. Salvatore Dello Iacono et al. "pH strip reader for beer samples based on image analysis". In:*2020 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor)*.2020, pp. 319–322.
2. Mohamed M. Elsenety et al. "Facile and highly precise pH-value estimation using commonpH paper based on machine learning techniques and supported mobile devices". In: *Scientific Reports* 12.1 (Dec. 2022).
3. Keshav Godha et al. "Human vaginal pH and microbiota: an update". In: *Gynecological Endocrinology* 34.6 (2018). PMID: 29271266, pp. 451–455.
4. Sung Deuk Kim, Youngmi Koo, and Yeoheung Yun. "A Smartphone-Based Automatic Mea- surement Method for Colorimetric pH Detection Using a Color Adaptation Algorithm". In: *Sensors* 17.7 (July 2017), p. 1604.
5. Yen-Pin Lin et al. "Vaginal pH Value for Clinical Diagnosis and Treatment of Common Vaginitis". In: *Diagnostics* 11.11 (2021).
6. BY Loh et al. "Automated Mobile pH Reader on a camera phone". In: *IAENG International Journal of Computer Science* 38.3 (2011), pp. 268–274.
7. Malene Risager Lykke et al. "Vaginal, Cervical and Uterine pH in Women with Normal and Abnormal Vaginal Microbiota". In: *Pathogens* 10.2 (Jan. 2021), p. 90.
8. Fabiana Nakano, Rogério Leão, and Sandro Esteves. "Insights into the role of cervical mucus and vaginal pH in unexplained infertility". In: *Medical Express* 2 (Mar. 2015).
9. Bianca Maria Pazzi, Dario Pistoia, and Giancarla Alberti. "RGB-Detector: A Smart, Low- Cost Device for Reading RGB Indexes of Microfluidic Paper-Based Analytical Devices". In: *Micromachines* 13.10 (Sept. 2022), p. 1585.

10. Subir Roy et al. “Improving Appropriate Use of Antifungal Medications: The Role of an Over-the-Counter Vaginal pH Self-Test Device”. In: *Infectious Diseases in Obstetrics and Gynecology* 11.4 (2003), pp. 209–216.
11. Ramon Sanchez-Iborra and Antonio F. Skarmeta. “TinyML-Enabled Frugal Smart Objects: Challenges and Opportunities”. In: *IEEE Circuits and Systems Magazine* 20.3 (2020), pp. 4–18.
12. Vinicius Aparecido Oliani Pedro da Silva et al. “Microfluidic paper-based device integrated withsmartphone for point-of-use colorimetric monitoring of water quality index”. In: *Measurement*164 (Nov. 2020), p. 108085.