An Effective Method to Detect Tonsillitis Using Machine Learning

Tamal Joyti Roy[1*] and Nobonita Saha[2]

[1*]Institute of Information and Communication Technology, Khulna University of Engineering & Technology, , Khulna , 9203, Bangladeh.

[2]Institute of Nutrition and Food Science , University of Dhaka, , Dhaka, 10000, , Bangladesh.

*Corresponding author(s). E-mail(s): tjroy13june@gmail.com; Contributing authors: nobonitasahaqueen@gmail.com;

Abstract

Objective: The purpose of this research was to predict the tonsillitis using machine learning algorithms. Increasing utilization of smartphones with sensor systems and machine learning capability promise better M-Healthcare services. Tonsillitis is an inflammation of your tonsils. Tonsillitis analysis requires contemporary technology.

Method: Different machine learning algorithms and frameworks used for evaluating the accuracy and performance.Artificial Neural Networks combined with picture processing and RGB color coding used for identify tonsillitis early and monitor prognosis at home. This study describes an innovative machine learning and smartphone-based optimization approach with a linked camera.

Results: Patients in remote locations, poor and impoverished countries may check, assess, and frequently do tonsillitis exams anywhere, anytime, and any place. This research proposes an unique method and machine learning approach to evaluate tonsillitis photos and diagnose infections with 90% accuracy for Random Forest and Decision Tree, .

Conclusion: In this research paper, we have introduced an advanced MHealth application on human health and monitoring systems. The use and technological advancement of smartphones has skyrocketed in the last decade. Now embedded sensors in smartphone devices help to assess physiological indicators and evaluate the health status. We have demonstrated that M-health can be effectively applied in the detection of tonsillitis by using smartphone devices and machine learning.

Keywords: Machine Learning, M-Health, Digital Health, Public Health

1 Introduction

M-Health is the practice of mobile phones and other wireless technologies in medical treatment. Mobile devices are most often used to educate customers about preventative healthcare. Illness surveillance, therapeutic assistance, epidemic outbreak monitoring, and chronic disease management also employ M-Health. In underdeveloped regions with a big population and substantial mobile phone usage, M-Health is popular.M-Health Alliance promotes M-Health in underdeveloped countries.M-Health to evaluate and upgrade a patient's assessment, treatment, surveying, and education about health conditions that end up escalating patient participation, boosting health outcomes, and restraining costs [1]. M-Health services have become now a growing concern for the applications such as monitoring oxygen saturation level of blood, heart rate, blood pressures, skin diseases, body temperatures, and many more [2] via smartphones. In ENT pathology, the ranking of tonsillitis is third in number [3]. Tonsillitis can occur at any age. Young children and school-going children generally aged 5 to 15 years are the most common risk group of tonsillitis and appear mainly in the winter season. There are three types of tonsillitis depending on symptoms and recovery period- acute, chronic, and recurrent [4]. The termination of the symptoms for acute tonsillitis is about 3 to 4 days but lasts about 2 weeks, caused mainly by viruses. A person with recurrent tonsillitis has several case histories of acute tonsillitis in a year and chronic tonsillitis results in progressive sore throat and malodorous breathing for a long period [5]. So our study objectives are-

- This article uses an algorithm-based visual system captured by smartphone cameras and artificial neural networks to identify tonsillitis using machine learning method
- We have proposed a new hybrid machine learning framework
- We anticipated the accuracy more than 90
- M-health Integration

2 Related Works

This research aimed to locate the Health Technology Assessment (HTA) reports on M-Health platforms and tried to investigate their integrity, regularity, and correctness to identify areas for improvement. A. W. Haidara et al also felt the importance of M-Health and sought HTA studies on M-Health systems and examined their integrity, regularity, and accuracy in suggesting improvement opportunities [6]. This study evaluated the usefulness, reliability, and validity of M-Health, its contribution to lifestyle and medical applications for medicare facilities. Practicality M-Health provides convenience, continued usage, observable acceptance, and user's ability to operate digital applications. User satisfaction, acceptance and recommendations influences app approval and Overdijkink, Sanne B., et al suggested the same idea in M-Healthcare applications [7]. Moreover, Ov Dou, Kaili, et al. aimed to establish and revise a conceptual framework to identify and explain factors influencing patients' acceptance of smartphone based healthcare services for chronic conditions [8]. M-Health (also can be termed as e-health) aims to ensure all patients get rational therapy and improve patient and sickness safety. A study of Zaman, Sojib Bin, et al. explained the e-Health clinical systems briefly [6] Mobile technology (M-Health) is utilized to support individuals with illness management, pleasurable encounters, and experiential growth in e-Health. According to a survey by Gorini, Alessandra, et al in 2018 addressed that "customers don't always use such technologies longterm". This paper examined the factors impacting M-Health uptake in health care [10]. The etiology, diagnosis, examination, and treatment of each condition are needed to be emphasized, along with guidance on urgent medication and referral criteria to an ENT unit. We'll synthesize fundamental ideas and research data to support these executive judgments. Bartlett, A., S. Bola, and R. Williams (2015) also emphasized on tonsillitis as a inflammation which can be life threatening when occurs chronic inflammations [11]. Tonsillectomy studies, methods for optimal outcomes, surgical pain alleviation strategies, and ways to lessen surgical hemorrhaging dangers, and cost research analysis are needed to be explored. Skevas, Theodoros, et al.(2010) conducted a study to assess the quality of life in adults with chronic tonsillitis and suggested the needs to research on this field [12]. Our study aims to provide a thorough, systematic approach to image processing and decision-making using sophisticated computational tools. Clinical image analysis methodologies and techniques were found, investigated, and discussed with an emphasis on computers by Gygli, Savina, et al.(2019).The solutions mentioned in that indicated how far we've come in reducing computer latency and the difficulties that remain [13]]. Image processing technologies handle a (output) picture based on various inputs. Images may be classed by the radioactive element or location using the investigated attribute, or whether they are made consciously or implicitly. Diagnostic imaging technologies, for example, employ x-ray retardation or ultrasound reflection data. Analog pictures may be transformed to digital images to appropriately portray the supplied picture. Dougherty, Geoff. (2009) presented an expandable, platform-independent image processing and visualization tool for cyber-connected health scientists [14]. MIPAV (Medical Image Processing, Analysis, and Visualization) offers online medical and graphical clinical image processing. Researchers and practitioners in remote places may easily communicate experimental data and statistics utilizing MIPAV's standard user experience and statistical approaches. The effective application of MIPAV was discussed in a study by McAuliffe, Matthew J., et al. (IEEE, 2001) [[15]. In this study, we focus to implement image processing in detecting tonsillitis which can address the appliance of M-Health. Our study

Table 1 Descriptive Data Analysis Result

Count	Mean	std	min	25%	50%	75%	max
55.00	44.527273	28.257975	1.00	20.500	45.00	68.000	97.00
55.00	163.660182	37.066781	95.67	132.830	161.33	195.65	230.33
55.00	67.042182	26.396500	3.33	53.000	64.67	84.000	137.00
55.00	69.479091	25.172502	22.00	56.165	68.33	83.335	143.33
55.00	0.709091	0.458368	0.00	0.000	1.00	1.000	1.00

will try to address the research gaps on this field. Our method is based on neural networks and image processing. We have collected more than a hundred images from Kaggle online database [16]. We have proposed a feature enrich system that could analyze tonsillitis. Our source data set consisted of data from different ages. All we needed was a smartphone that capable of taking pictures with a flashlight. Our working procedure consisted of neural network implementation. Our algorithm was trained for processing the data based on RGB color code. And it has been found significant evidence that the RED color plays a role in detecting tonsillitis.

3 Proposed Methodology

Our method was based on machine learning and image processing. We have collected more than a hundred images from [16] . We have proposed a feature enrich system that could analyze tonsillitis. Our source data set consisted of data from different ages. All we needed is a smartphone that capable of taking pictures with a flashlight. Figure 1 showing our sample data set. Our working procedure consisted of neural network implementation. Our algorithm was trained for processing the data based on RGB color code. We have given the value to if the value of R is larger than 150 disease and if the value of R is less than 150 then it's not infectious. The Value of G and B constantly vary throughout the whole experiment. Figure 2 showing our whole experimental analysis in a diagram. We performed our coding in python 3.8 with the windows cmd console. We have analyzed with bare eyes and compared the data with the analysis, which came out to be more perfect with the coding outcomes.

The neural network system was able to predict the tonsillitis. There were ten nodes in the input layers, and the hidden layers were twenty nodes. The output showed as Tonsillitis or not Tonsillitis. A quad-core processor was used with 8 GB DDR4 ram for the experimental analysis. The whole process took less than two minutes. Table 1 showing the correlation of the features selections.There was a signi cant correlation with Red color, which was our prediction based upon. Table 2 showing the results of the one-sample test. That to verify our hypothesis. And we found significant evidences that the RED color plays a role in detecting tonsillitis.

Multi layer Perceptron : The feed forward neural network now now includes the multi layer perceptron, which is sometimes referred to as MLP in certain circles. There are three distinct types of layers that come together to

Fig. 1 Sample Dataset

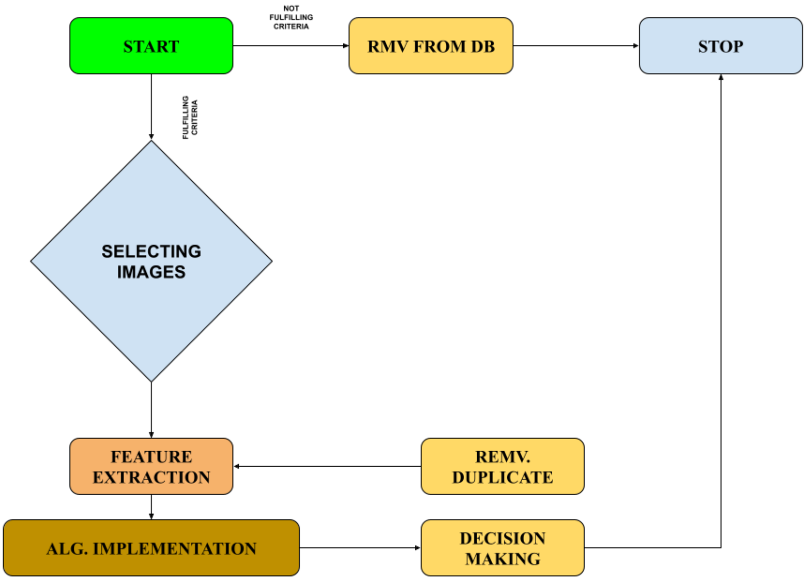

Fig. 2 Experimental Flowchart

Table 2 Different Types of Data

Sl No	Columns	Non-Null Count	Dtype
1	ImageName	55 not-null	int64
2	Red	55 not-null	oat64
3	Green	55 not-null	oat64
4	Blue	55 non-null	oat64
5	TonsilitisYesorNo	55 not-null	int64

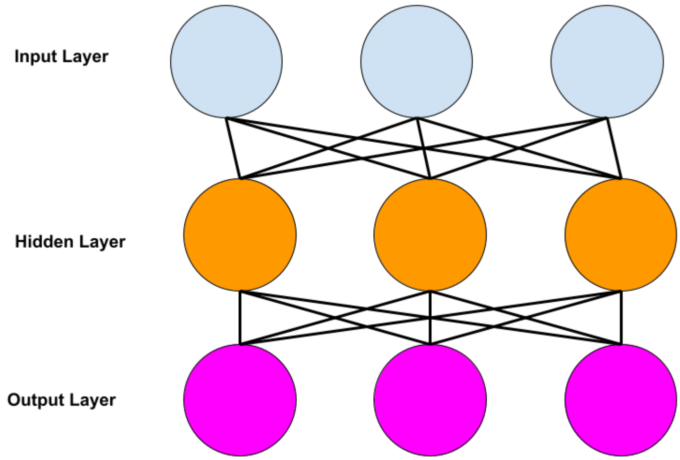

Fig. 3 Typical MPL Architecture

form this structure. These layers are referred to as the input layer, the output layer, and the hidden layer. The input layer receives the signal that is going to be processed once it has been passed on to it.
The output layer is in charge of nishing the essential tasks, which may include prediction and classi cation, and it is accountable for doing so. In a multilayer perceptron (MLP), the true computational engine is made up of an arbitrary number of hidden layers that are located in the middle of the visible layer layer. A multilayer perceptron (MLP) works in the same way as a feed forward network does. Data flows forward from the input layer to the output layer. The neurons that make up the MLP are instructed through a method known as back propagation learning. Because of its architecture, which allows them to approximate any continuous function, MLPs can handle issues that cannot be linearly divided into many sections. The key application cases for machine learning patterns include pattern classification, identification, prediction, and

approximation.
MLPs are a kind of neural network model that may operate as universal approximators, which means that they are able to provide a close approximation of any continuous function . For instance, you may make use of them as SEE models. Perceptions are the neuronal building blocks that make up MLPs. First, the overall structure of a perceptron will be discussed, and then we will go on to discussing the general structure of MLPs. A perceptron has n characteristics that it takes in as input (x = x1, x2,..., xn), and each of these features has a weight that corresponds to it. All of the input characteristics have to be numbers. In order to employ a perceptron, then, non-numerical input properties will need to be transformed into numerical ones. For instance, a category feature with p potential values may be transformed into p input features expressing the presence or absence of these values. These values are represented by "yes" or "no."These are referred to as dummy variables. For example, if the input feature "development type" accepts the values "new development," "enhancement," or "re-development," it may be possible to replace it with three dummy variables named "new development," "enhancement," and "redevelopment," each of which takes the value 1 if the corresponding value is present and 0 if it is not. These factors would determine whether or not the appropriate value is present.
Random Forest: Random Forest is a well-known machine learning technique that falls within the more general field of supervised learning. Classification and regression are two examples of machine learning tasks that might benefit from its use. It is based on the concept of ensemble learning, which refers to the process of merging several classifiers to solve a difficult problem and improve the model's functionality. "Random Forest" is a classifier that "contains a number of decision trees on various subsets of the provided dataset and takes the average to increase the prediction accuracy of that dataset," as the name implies. Random Forest is a strategy that "averages the forecasting accuracy of that dataset." Random Forest is a technique that "takes the average to improve the predictive accuracy of that dataset." The tree does not rely on a single decision tree; rather, it takes into account the forecast from each tree in the forest and calculates the final output depending on which tree's prediction earned the most votes. The greater the number of trees in the forest, the higher the level of accuracy attained and the avoidance of errors.

Decision Tree : Although the decision tree technique to supervised learning may be used to solve problems requiring both classification and regression, it is most commonly employed to solve classification difficulties. It is a tree-based classifier, with core nodes reflecting dataset properties, branches representing decision rules, and each leaf node providing the classification's result.
KNN: K Nearest Neighbor (KNN) is one of the most e ective machine learning algorithms available today. It is very user-friendly, straightforward, and exible. KNN is used in a wide number of applications including but not limited to the following: handwriting recognition, image recognition, video recognition,

and the elds of nance and healthcare. The process of determining a customer's credit rating is one that is carried out by nancial institutions. When it comes time to hand out loans, nancial institutions will make a determination as to whether or not the loan poses a risk. In the eld of political science, prospective voters are divided into two categories: those who will vote and those who will not vote. The KNN technique is used to issues involving both classi cation and regression. KNN algorithm that uses an approach based on feature similarity.
Logistic Regression: Logistic regression is a technique for estimating the probability of a discrete outcome given an input variable. The input variable is the most important component in this method. The most common type of logistic regression models a binary outcome, which refers to anything with only two potential values, such as true or false, yes or no, and so on. When a scenario has more than two discrete outcomes, multinomial logistic regression can be utilized to depict the issue.Logistic regression is a good way of analysis for classi cation issues, where you are attempting to decide whether a new sample ts best into a category. This kind of problem arises when you have a large amount of data to analyze. Logical regression is a helpful analytical approach that may be used since many parts of cyber security include classi cation di culties. One example of this would be the detection of attacks.

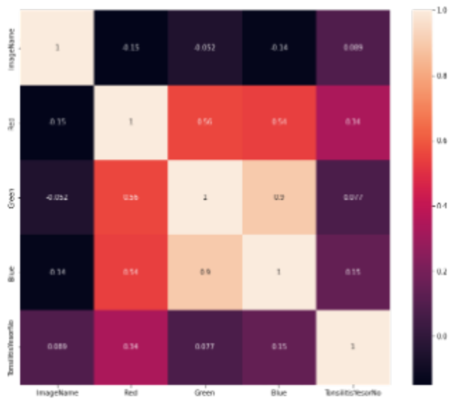

Fig. 4 Heat-Map

SVM: The Support Vector Machine, abbreviated SVM, is one of the most extensively used supervised learning algorithms. It may be applied to classification and regression problems. However, its principal use is in the field of machine learning, specifically for classification difficulties. The Support Vector Machine (SVM) approach is used to find the best line or decision boundary to split an n-dimensional space into classes. In the future, we will be able to simply insert any new data points in the proper category. This ideal decision boundary is referred to as a hyperplane..
Stochastic Gradient Decent: One of the most well-known and commonly

used Machine Learning methods is the stochastic gradient descent technique. It is well known for serving as the foundation for neural networks. Simply described, a "gradient" is the slope or inclination of a surface..Therefore, gradient descent refers to the process of going down a slope in order to reach the lowest point on that surface. "Gradient descent is an iterative process that begins at a random point on a function and moves down the slope of that function in stages until it reaches the lowest point of that function," "Gradient descent begins at a random point on a function."
Linear SVC: The purpose of a Linear SVC, also known as a Support Vector Classi er, is to conform to the information that we provide it and then produce a hyperplane that is the "best t" for dividing or classifying our data. Once we have the hyperplane, the next step is to enter some characteristics into our classi er to determine what the "predicted" class will be. This particular method is thus very good for our applications, despite the fact that we may utilize it for a wide variety of settings.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(fi - yi)^2 \tag{1}$$

To determine which branch is the best choice for forest, this algorithm estimates the distance between each node and the expected actual value.

$$E\left(S\right) = \sum_{i=1}^{c} -p_i log_2 \tag{2}$$

Where S stands for "Current state" and Pi for "Probability of an Event I in State S" or "Percentage of Class I in a State S Node"

$$K(N) = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \tag{3}$$

The multi - linear model's core formula is

$$E(y)) = \alpha + \beta x1 + \lambda x2 \tag{4}$$

Naive Bayes: The Naive Bayes classi ers are a group of classi cation algorithms that may be traced back to Bayes' Theorem as their point of origin. It is not a single algorithm but rather a family of algorithms that all share a similar assumption, which is that every pair of qualities being categorized is independent of each other. This premise is what makes it an algorithm family rather than a single algorithm. This is due to the fact that it is not just one algorithm but rather a family of algorithms working together.

4 Results

In Figure 6, we have shown the image analysis of tonsillitis.We have seen the 9 machine learning models predictions for tonsillitis. The neural link helped to distinguish tonsillitis by looping through the neural link nodes and made two last decisions. Figure 4 showing the predicting accuracy of tonsillitis. Where the accuracy of prediction varies from 0.4 to 0.9 that means we were able to achieve the accuracy of more than 90 percent with random forest and decision tree. All the analysis were tested for cross check on a computer and a mobile.

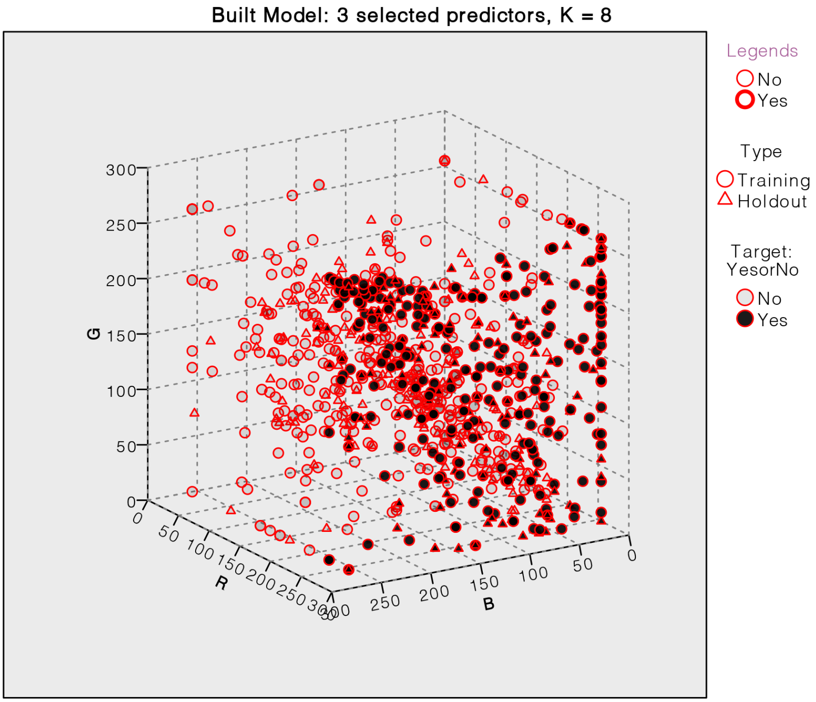

Fig. 5 KNN classification model (k=8))

Training data set was seventy percent in number and testing thirty percent. Image one to seventy and image seventy one to hundred. The Python virtual environment was created to simulate the process. The error rate varies from 0.6 to 0.1 percent. That was low and good for the experiment. The energy consumption of the experimental analysis was negotiable. We performed our test in windows OS and Linux OS. The Linux OS system made the calculation much quicker than the Windows platform. So our experiment was able to predict tonsillitis using image processing and a neural network system.Even

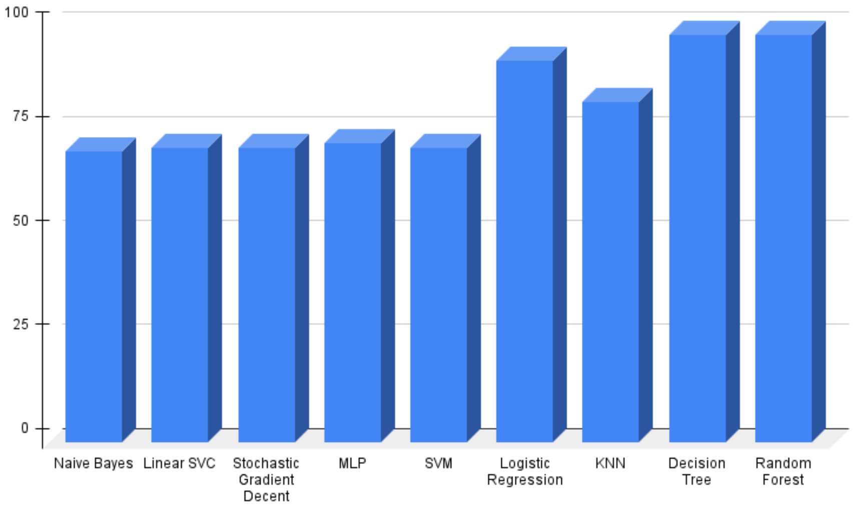

Fig. 6 Different ML model accuracy

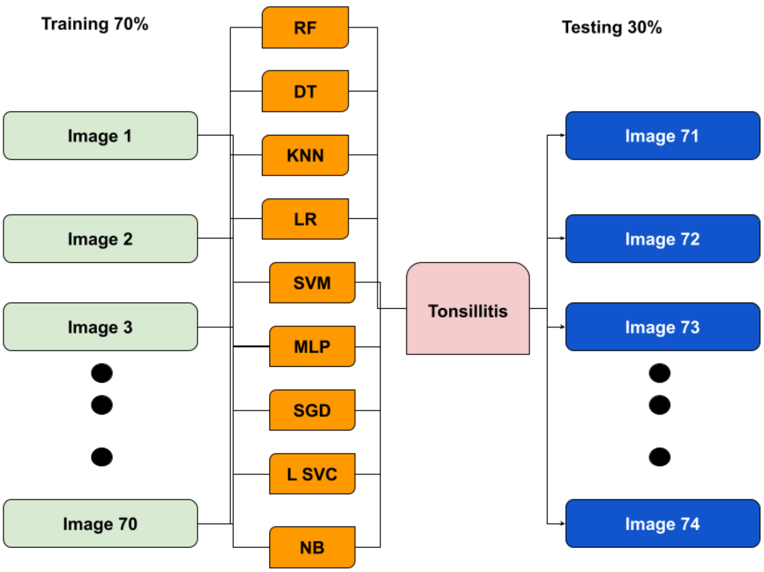

Fig. 7 Training and Testing Phase

though these strategies were prompted by particular methodologies, we believe that they are nevertheless broadly applicable and have the potential to progress the area of informed machine learning as a whole. So we were able to captured random images from tonsillitis su erer and able to identify the condition with more than ninety percent accuracy classi cation . Our hybrid machine learning framework has been working. Classify accuracy was 98% (Random Forest and Decision Tree). And we were able to integrate with M-health technology.

5 Conclusion

In this research paper, we have introduced an advanced MHealth application on human health and monitoring systems. The use and technological advancement of smartphones has skyrocketed in the last decade. Now embedded sensors in smartphone devices help to assess physiological indicators and evaluate the health status. We have demonstrated that M-health can be e ectively applied in the detection of tonsillitis by using smartphone devices and machine learning. In particular, our work highlights that machine learning with neural networks can di erentiate amongst normal images with infected images of tonsils which can easily show us the result if there is any in ammation or not. We have been able to create an algorithm-based imaging system using RGB color code in images recorded by smartphone devices. The coding system has been approved to analyze the tonsils and thus we can obtain color code signatures for each image to identify tonsillitis. Previously for early detection of tonsillitis, individuals with symptoms of tonsillitis had to visit doctors for physical examination, and numerous medical tests had to be performed to diagnose tonsillitis with high accuracy. The system developed here would be extremely bene cial for early or pre-detection of tonsillitis, as well as the continuous monitoring at home and remote medical care with minimal cost. Our method performs an equivalent or preferable response to otorhinolaryngologists in terms of identifying tonsillitis as tonsillitis can be diagnosed easily by physical examination. The results shown here demonstrate that smartphone-based imaging analysis linked with neural networking systems has a great potential as MHealth to diagnose tonsillitis. It has already been understood that smartphones with machine learning expert systems and improved neural networks will be worthwhile to public health investigators in revealing tonsillitis, and most likely, several forms of human pathologies in medical trials. This enduring implementation will proceed to test the tonsillitis detection in global health settings and we are optimistic about forthcoming innovations that will strengthen diagnostic precessions, speci city, and selectivity.

References

[1] M. Fiordelli, N. Diviani, and P. J. Schulz, "Mapping mHealth Research: A Decade of Evolution," Journal of Medical Internet Research, vol. 15, no. 5, p. e95, May 2013, doi: 10.2196/jmir.2430.

[2] A. Bourouis, A. Zerdazi, M. Feham, and A. Bouchachia, “M-Health: Skin Disease Analysis System Using Smartphone’s Camera,” Procedia Computer Science, vol. 19, pp. 1116–1120, 2013, doi: 10.1016/j.procs.2013.06.157.

[3] A. W. Haidara, Y. Sidibé, D. Samaké, and A. A. Mohamed, “Tonsillitis and Their Complications: Epidemiological, Clinical and Therapeutic Profiles,” Scientific Research Publishing, Jan. 01, 2019.

[4] Windfuhr, J. P., Toepfner, N., Steffen, G., Waldfahrer, F., & Berner, R. (2016). Clinical practice guideline: tonsillitis I. Diagnostics and nonsurgical management. European Archives of Oto-Rhino-Laryngology, 273(4), 973-987.

[5] A. Pietrangelo, “Everything You Need to Know About Tonsillitis,” Healthline Media, Apr. 18, 2016. https://www.healthline.com/health/tonsillitissymptoms.

[6] A. W. Haidara et al., “Tonsillitis and Their Complications: Epidemiological, Clinical and Therapeutic Profiles,” International Journal of Otolaryngology and Head & Neck Surgery, vol. 8, no. 3, pp. 98–105, Apr. 2019, doi: 10.4236/ijohns.2019.83011.

[7] Overdijkink SB, Velu AV, Rosman AN, van Beukering MD, Kok M, Steegers-Theunissen RP The Usability and Effectiveness of Mobile Health Technology–Based Lifestyle and Medical Intervention Apps Supporting Health Care During Pregnancy: Systematic Review JMIR Mhealth Uhealth 2018;6(4):e109

[8] Dou K, Yu P, Deng N, Liu F, Guan Y, Li Z, Ji Y, Du N, Lu X, Duan H Patients’ Acceptance of Smartphone Health Technology for Chronic Disease Management: A Theoretical Model and Empirical Test JMIR Mhealth Uhealth 2017;5(12):e177 doi: 10.2196/mhealth.7886

[9] Zaman SB, Hossain N, Ahammed S, Ahmed Z. Contexts and Opportunities of e-Health Technology in Medical Care. J Med Res Innov. 2017;1(2):AV1-AV4. DOI: 10.15419/jmri.62

[10] Gorini A, Mazzocco K, Triberti S, Sebri V, Savioni L and Pravettoni G (2018) A P5 Approach to m-Health: Design Suggestions for Advanced Mobile Health Technology. Front. Psychol. 9:2066. doi: 10.3389/fpsyg.2018.02066

[11] Bartlett, A., Bola, S., & Williams, R. (2015). Acute tonsillitis and its complications: an overview. Journal of the Royal Naval Medical Service, 101(1).

[12] T. Skevas, C. Klingmann, S. Sertel, P. K. Plinkert, and I. Baumann, "Measuring Quality of Life in Adult Patients with Chronic Tonsillitis," Open Otorhinolaryngol. J., vol. 4, no. 1, pp. 34–46, Jun. 2014, doi: 10.2174/1874428101004010034.

[13] Gulo, C.A.S.J., Sementille, A.C. & Tavares, J.M.R.S. Techniques of medical image processing and analysis accelerated by high-performance computing: a systematic literature review. J Real-Time Image Proc 16, 1891–1908 (2019). https://doi.org/10.1007/s11554-017-0734-z

[14] "Digital Image Processing for Medical Applications - Geoff Dougherty - Google Books." https://books.google.com.bd/books?hl=en&lr=&id=ARtSAl-A-lsC (accessed Jul. 31, 2021).

[15] M. J. McAuliffe, F. M. Lalonde, D. McGarry, W. Gandler, K. Csaky and B. L. Trus, "Medical Image Processing, Analysis and Visualization in clinical research," Proceedings 14th IEEE Symposium on Computer-Based Medical Systems. CBMS 2001, 2001, pp. 381-386, doi: 10.1109/CBMS.2001.941749.

[16] "Different Phases of Tonsilitis," https://www.kaggle.com/datasets/anindamohanta/different-phases-of-tonsilitis (accessed Aug. 20, 2022).