








ISSN: 2617-6548

URL: www.ijirss.com



A Mobile Computer-Aided Diagnosis of Neonatal Hyperbilirubinemia using Digital Image Processing and Machine Learning Techniques

 Supaporn Dissaneevate¹,  Thakerng Wongsirichot^{2*}, Pittaya Siriwat³, Nutchaya Jintanapanya⁴, Uakarn Boonyakarn⁵,
 Waricha Janjindamai⁶,  Anucha Thatrimontrichai⁷,  Gunlawadee Maneenil⁸

^{1,4,6,7,8}Division of Neonatology, Department of Pediatrics, Faculty of Medicine, Prince of Songkla University, Songkhla, Thailand.

^{2,3,5}Division of Computational Science, Faculty of Science, Prince of Songkla University, Songkhla, Thailand.

*Corresponding author: Thakerng Wongsirichot (Email: thakerng.w@psu.ac.th)

Abstract

Neonatal Hyperbilirubinemia, or jaundice, is a harmful disease found in newborns, a symptom of which is the yellowish discoloration of the skin. Visual examination is most frequently used for screening of Hyperbilirubinemia in neonates, however, blood specimen collection is the gold standard to identify the disease and its severity. We propose a Mobile Computer-Aided Diagnosis (mCADx) tool to identify the Neonatal Hyperbilirubinemia symptom using advanced digital image processing and data mining techniques. The mCADx was developed in a cross-platform environment. The mCADx works with smart devices run on either iOS or Android operating systems. With ethical committee approval, we collected and studied image data of 178 infant subjects with different jaundice severity levels. The severity of the disease was examined from blood test results, which were annotated by medical specialists. Data mining techniques included Decision Trees, k Nearest Neighbor, and the Conventional Neural Network was investigated in the dataset. An in-depth comparison between techniques was performed and discussed. The classification results in CNN gained the highest accuracy at 0.8099, 0.9251, 0.8086. This novel work can assist in identifying Neonatal Hyperbilirubinemia in newborns after discharging from the hospital. Reoccurring Neonatal Hyperbilirubinemia can be found with minimum awareness of parents. Limitations and future works were discussed in this work.

Keywords: Neonatal Hyperbilirubinemia, Jaundice, Skin color, Digital image processing, Machine learning, mCADx.

DOI: 10.53894/ijirss.v5i1.334

Funding: This study received no specific financial support.

History: Received: 8 November 2021/**Revised:** 20 December 2021/**Accepted:** 5 January 2022/**Published:** 24 January 2022

Copyright: © 2022 by the authors. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Authors' Contributions: All authors contributed equally to the conception and design of the study.

Competing Interests: The authors declare that they have no competing interests.

Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study was reported; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained.

Ethical: This study follows all ethical practices during writing.

Publisher: Innovative Research Publishing

1. Introduction

Neonatal Hyperbilirubinemia is one of the most harmful diseases found in infants. Approximately 8-11% of infants develop symptoms within the first seven days after birth. Clinically, the Hyperbilirubinemia symptom is detected through the Total Serum Bilirubin (TSB) test. If an infant has TSB above the 95th percentile for age during the first seven days of life, Hyperbilirubinemia is considered. Apart from the first seven days, approximately 60%-80% of healthy infants could show signs of idiopathic neonatal jaundice, and after birth, about 20-50% [1]. In general, the symptom is caused by the breakdown of the red blood cells. At the same time, bilirubin is continuously increasing. The symptom is a severe condition and can cause severe adverse effects on infants. Prolonged symptoms may cause further complications. Reoccurring Neonatal Hyperbilirubinemia can be found in many infants after discharging from hospitals with or without parental awareness.

2. Background

Neonatal Hyperbilirubinemia, or jaundice, is the yellowish discoloration of the skin, conjunctiva, and mucous membrane. It can be observed when serum bilirubin level exceeds 5 mg/dL [2]. Generally, it is the consequence of increased bilirubin production from the destruction of red blood cells. Additionally, insufficient hepatic conjugation and increased enterohepatic bilirubin reuptake causes the reduction of bilirubin elimination [3-5]. Newborns commonly develop some degree of physiological jaundice that usually appears two to four days after birth and resolves spontaneously after one to two weeks [4]. However, some cases develop potentially hazardous bilirubin levels that lead to acute bilirubin encephalopathy, and in most severe cases, kernicterus can be found. Kernicterus, a form of permanent brain damage, causes death or long-term neurodevelopmental disability. It includes athetoid cerebral palsy, sensorineural deafness, upward gaze paresis, and dental enamel hypoplasia [4, 6, 7]. Consequently, early detection is essential for preventing the complications associated with neonatal Hyperbilirubinemia, and prompt phototherapy or exchange transfusion can be performed as an initial treatment for severe cases.

Visual examination by Kramer's scale is the most frequently used for screening of Hyperbilirubinemia in neonates. Jaundice is usually seen first in the stages, progressing cephalocaudally to the lower body parts, which can be detected by blanching the skin with digital pressure on the bony prominence of the forehead, mid-sternum, knee, and ankle. However, this visual assessment can be subjective and inaccurate, often confounded by skin color, hemoglobin, ambient light, and observer experience [5, 8]. Blood specimen collection is the gold standard to measure the bilirubin level by counting total serum bilirubin concentration. However, invasive blood sampling in neonates causes pain and distress. In addition, blood loss incorporated with this technique can increase the risk of infection [9]. Therefore, researchers have developed non-invasive diagnoses to detect neonatal hyperbilirubinemia symptoms such as transcutaneous bilirubinometer [5]. However, with the newly developed non-invasive diagnose technology, the high cost is troublesome in a resource-limited setting.

In addition to medical insights, Digital Image Processing and Data Mining Techniques have been included in this research work. Data mining techniques can be classified into various groups based on the purposes of the research. The goals include trend analysis, prediction, classification, clustering, trend analysis, and identifying unforeseen data patterns in particular problems. Selected data mining techniques are mainly based on the studied data and research experiences, and a supervised learning method is usually applied in classification problems. Specifically, a selected dataset is pre-labelled by a reliable source or specialist.

A data mining project standard was initiated as a common framework, designed to follow the Cross-Industry Standard Process for Data Mining (CRISP-DM). CRISP-DM includes six steps: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment [10, 11]. Fundamentally, the CRISP-DM begins from understanding the business context or research area of the problem without concerning about data. Data understanding is the step in which data structures are studied to learn about the data inputs from various available sources. After the data inputs are analyzed, the data are prepared for modeling. The data preparation process is one of the essential stages. High-quality data with minimized errors should be the primary concern. Multiple machine learning models can be used in the modeling stage such as Artificial Neural Network (ANN), Decision Tree (DT), Random Forest (RF), k Nearest Neighbor (k NN), clustering, etc. Recently, the Deep Learning technique has been highly used and claimed to be one of the most powerful techniques. However, common trade-offs are longer processing time and high-performance computers required. Various machine learning models are still active and able to solve many problems. For example, Decision Tree (DT) uses the concept of recursive selection. It identifies the best appropriate data fields to separate data chunks. The tree's leaf nodes grow until a predefined stopping criterion is met. A gain ratio calculation is used in the splitting procedure by automatically assigning weight values to each attribute. The evolution of DT leads to a new machine learning model called Random Forest (RF). RF is known as an ensemble learning method that is primarily used in classification problems. Fundamentally, it is built from multiple decision trees at the training time. In the clustering model, the k Nearest Neighbors (k NN) is used in many research problems. k NN, a non-parametric method, assigns weight values to the contributions of the neighbors from a reference point. The nearer neighbors are identified into the same cluster or group as the reference data point. In other words, the closer neighbors highly contribute to the average than more distant ones. Another machine learning model is widely applied in many areas. Artificial Neural Network (ANN) is designed based on organized layers. Each layer contains several interconnected call nodes. ANN works on a selected activation function. It holds the key to identifying and interacting between nodes in a layer.

Over the years, researchers collected and used medical images such as skin photos, x-ray, and MRI images for studying symptom and disease classifications. In disease classification, image processing techniques play an essential role. A vital background theory in image processing techniques is the color models. The RGB color model processes an image as a composite of three grayscale images that correspond to the intensity of Red, Green, and Blue light, respectively. The R, G, and B grayscale images can be recombined into a single image, which humans can perceive as a color image. The HSV is

derived from Hue, Saturation, and Intensity. The HSV color model claims to be a more consistent color model to human visual perception. Hue refers to a particular wavelength of light that corresponds to the actual perceived colors of an image. Saturation and intensity represent the density of the hue that reaches human eyes and the brightness of an image, respectively. The LCH or HCL model refers to Hue, Chroma, and Luminance, which uses the HSV concept. It retains the human perception of color by using three parameters, H, C, and L. It limits the bias from the variation of saturation, possibly found in the HSV color model [12-14]. The International Commission firstly defined the LAB or CIELAB color model on Illumination (CIE) in 1976. It deploys color based on three values: L^* for the lightness ranging from black (0) to white (100), a^* from green to red, and b^* from blue to yellow. The LAB color model is one of the closest approximations to human vision. It employs the L component to match the human perception of lightness [15].

Many research works have recently utilized machine learning and image processing techniques to analyze medical images for screening tests. BiliCam is a similar research work compared with our work, using a smartphone and a predefined low-cost color calibration card to detect and monitor Neonatal Hyperbilirubinemia from an infant's skin color. This research work gained 85% accuracy compared with the gold standard blood test for 100 newborns. However, there are some limitations that the researchers have mentioned. Firstly, the work has been designed and tested with only one type of device, the Apple iPhone 4s. Secondly, it relies on the low-cost color calibration card printed from ordinary printers. Therefore, it can cause a deviation of colors in various printers. Thirdly, the limitation of newborn subjects was discussed in terms of population and race [16]. In 2019, a non-invasive screening device for Neonatal Hyperbilirubinemia was introduced. The researchers proposed an instrument for detecting jaundice that generates a light beam on infants' nail beds through optical fibers. The Bland-Altman method was used to classify the TSB test and the proposed method, which showed a promising result [17].

3. Material and Method

We propose a cross-platform Mobile Computer-Aided Diagnosis (mCADx) to screen Neonatal Hyperbilirubinemia's symptoms using image processing and data mining techniques. The main goal of the developed mCADx was to be used as a screening tool for classifying Neonatal Hyperbilirubinemia symptoms over the infant skin photos.

Currently, public Neonatal Hyperbilirubinemia datasets with skin photos are rare. We collected a new dataset containing 178 infant skin photos from the newborn department at the Songklanagarind Hospital, Thailand. The Office of Human Research Ethics Committee (HREC) at the Songklanagarind Hospital, Thailand, formally approved collecting infant skin photos but restricted data distribution publicly, with the exception of this research. Therefore, the procedure was conducted with the highest standard regulated by the HREC. Photos were taken with the same standard procedure and setting. In addition, the procedure was performed in a room with similar light exposure to minimize external factors that can affect the classification model. Firstly, an infant was primarily placed in a supine position, in which the infant's shoulders and arms reside in the rest position beside its body. The infant's hands are usually fixed under the buttocks or placed next to its body. Due to unexpected and unintentional infant movements, some were not entirely in the supine position. Figure 1 shows the supine position of an infant during the procedure.

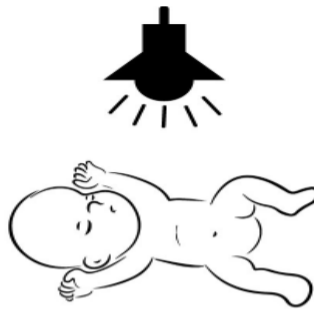


Figure 1.
Supine position.

The data collection procedure began once the infant was in the supine position. Three images were taken of each subject using three unique cameras, including a DSLR camera (Nikon D5100) and two mobile phone cameras. The used mobile phones were the iPhone 7 (iOS) and the Oneplus 5 (Android). Specifically, the Nikon D5100 was a digital camera that had the best resolution in our experiment at 16.2 megapixels. The iPhone 7 had a 4.70 inch (750x1334) display with 7MP resolution. The Oneplus 5 Android phone was equipped with 16 MP resolution with a 1080x1920 pixel display. All selected cameras were calibrated and their lights and flashlights disabled. The cameras were held vertically above the center area of the abdomen by approximately one meter. Photos were taken horizontally above the navel of each subject. As shown in previous medical evidence, the yellowness spread from head to toe, therefore, the body's center is an excellent position to have a photograph. Figure 2 shows the photographic area for our study.

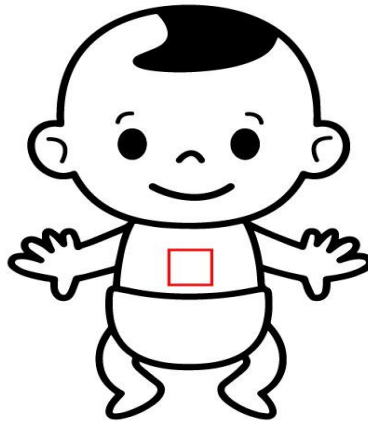


Figure 2. Photographic area.

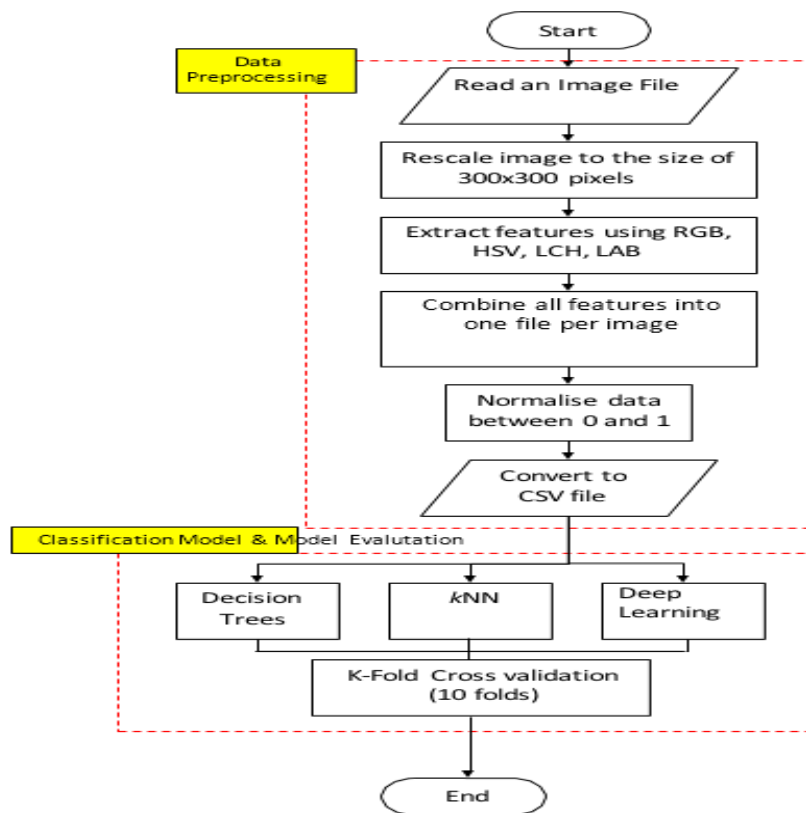


Figure 3. Experimental setting.

We designed our experimental setting to identify the best classification model. Substantially, a selected classification model was the core engine for our mCADx. Figure 3 shows the experimental setting for our study. It consists of two main stages: data preprocessing, and classification model and model evaluation.

A. Data Preprocessing

According to Figure 3, collected images were rescaled to the size of 300x300 pixels. Each image was processed using four different color models, including RGB, HSV, LCH, LAB. Processed features from the color models formed a dataset for the research problem. In addition, we classified the subjects into three categories based on the TcB level in each subject's blood test result. The categories included Severe (S) ($TcB \leq 10$), Medium (M) ($15 \leq TcB < 10$), and Low (L) ($20 \leq TcB < 15$). The TcB level of each subject was collected from the gold standard blood test and annotation specialists. There were 178 subjects in total for the research work. Figure 4 shows the distribution of the TcB level and the gestation age of Neonatal Jaundiced infants. Table 1 shows the distribution of the number of issues according to the TcB Level. Subsequently, statistical normalization was conducted. All data values were rescaled between 0 and 1. The statistical normalization compromised the data scale imperfection due to the substantial value differences between each feature. After the normalization, it was converted to a CSV file for further analysis.

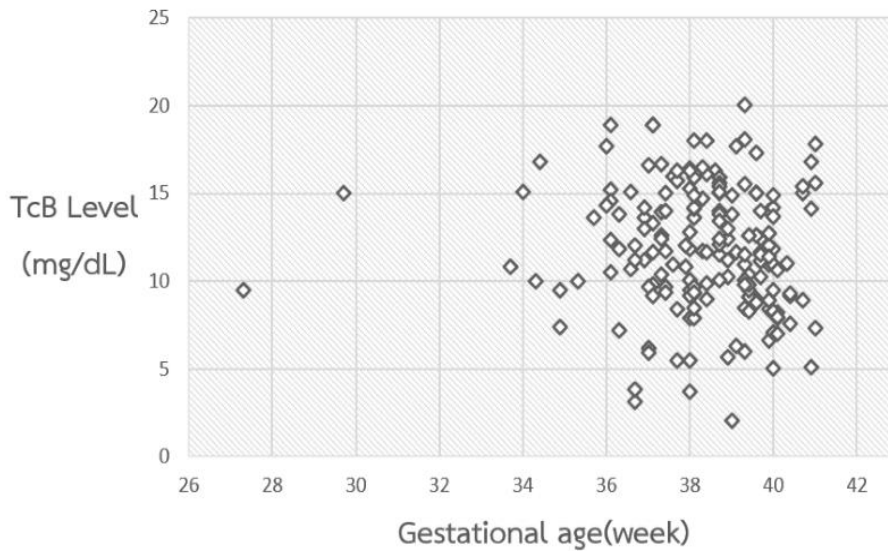


Figure 4.
The distribution of the TcB level, and the gestation age of Neonatal Jaundiced infants.

Table 1.
The TcB level.

Level	TcB Level	Count
Severe (S)	$TcB \leq 10$	57
Medium (M)	$15 \leq TcB < 10$	78
Low (L)	$20 \leq TcB < 15$	43

B. Classification Model and Model Evaluation

Decision Trees (DT), *k* Nearest Neighbors (*k*NN), and Convolutional Neural Networks (CNN) were selected for this research work. DT classified the input features from selected color models to the appropriate S, M, and L classes, which were the leaf nodes of this classification model. The process mimics the concept of conditional selection or if-then conditions in mathematics. In this research work, the second selected classification model, *k*NN, the preset *k* was 3, 4, 5, respectively. It worked by calculating the closest distance between test features and train features. The nearest value with the sample based on the *k* values was considered in the same class. The last classification model was CNN, which imitated the human brain's neural network, increasing unlimited processors. Generally, the preset neural network for this research work consisted of three layers. First, the input layer operated the input data into the algorithm. Next, the hidden layer utilized the data from the input layer to find a model or mathematical process related to the final layer. Finally, the output layer was the result of the algorithm related to the input layer [11].

A detailed comparison of these classification models was performed, using the sci-kit-learn library package in Python for our implementation. Training sets and testing sets were randomly selected in the model evaluation stage. In addition, the 10-fold cross-validation splits our data into ten different subsets. We used 9 (*k*-1) data subsets to train the model and the remaining subset (or the last fold) as a test set. Therefore, there was no data that has not been tested in this environment setting. Figure 5 shows the process of *k*-fold cross-validation [11].

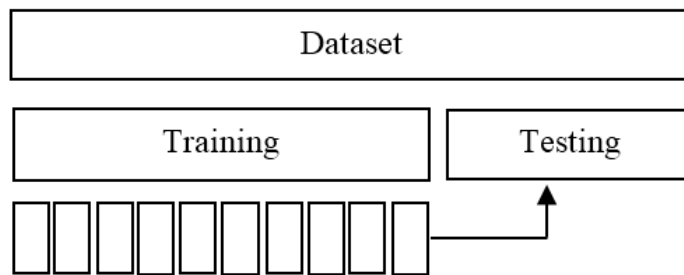


Figure 5.
k-fold cross-validation.

For model evaluation, performance measures have been selected. The selected measurements include Accuracy, Precision, Recall, Specificity, and F-Measure. The accuracy is a ratio of correctly predicted jaundice subjects to the total jaundice subjects. Precision represents the ratio of correctly predicted positive jaundice subjects to the total predicted positive jaundice subjects. Recall shows the ratio of correctly predicted positive jaundice subjects to all jaundice subjects in an actual class. Finally, the F-Measure mimics the weighted average of Precision and Recall. These selected measurements were

computed from incremental counts of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) in confusion matrices tableau.

4. Result

Tables 2, 3, and 4 show the classification results from three different cameras, iOS, Android, and DSLR, in three classification models, DT, *k*NN, and CNN. The maximum values are marked with *, and the second-highest values are marked with #. In terms of overall classification results, CNN performed best in all cameras. It achieved F-Measure at 0.8856, 0.8748, and 0.8848. Thus, CNN is a robust classification model claimed in many research works, however, it usually requires higher computing performance and is time-consuming [9]. In our case, the developed mCADx operated on limited computing power. Moreover, the timely manner was a primary concern for the screening test. Therefore, the second-best scenario was qualified to take into account. According to the F-Measure, DT performed promisingly in two cameras, iOS, and Android. In the DSLR, DT gained the third prize after the *k*NN (*k*=3). Theoretically, DT is a tree structure using reasonable conditions to separate nodes while consuming minimum computing performance [11], thus, it is suitable to be embedded into our developed mCADx.

Table 2.
Overall classification results in iOS camera.

Perf. Measure	Decision Trees		<i>k</i> Nearest Neighbor						CNN	
	Train	Test	Train (k=3)	Test (k=3)	Train (k=4)	Test (k=4)	Train (k=5)	Test (k=5)	Train	Test
Accuracy	0.7859	0.6667	0.7985	0.6852 [#]	0.6988	0.6481	0.6685	0.6296	0.9784	0.9688 [*]
Precision	0.5895	0.5758	0.5958	0.6158	0.6263	0.6285	0.5985 [#]	0.5394	0.7163	0.7289 [*]
Recall	0.5488	0.5896 [#]	0.4851	0.5985	0.6142	0.5987	0.5645	0.5298	0.6869	0.7826 [*]
Specificity	0.9185	0.8158 [#]	0.9236	0.7159	0.7358	0.8036	0.8144	0.6998	0.9356	0.9436 [*]
F-measure	0.8936	0.8269 [#]	0.8905	0.8658	0.8224	0.7436	0.7998	0.7936	0.8968	0.8856 [*]

Note: The maximum values are marked with *, and the second-highest values are marked with #.

Table 3.
Overall classification results in Android camera.

Perf. Measure	Decision Trees		<i>k</i> Nearest Neighbor						CNN	
	Train	Test	Train (k=3)	Test (k=3)	Train (k=4)	Test (k=4)	Train (k=5)	Test (k=5)	Train	Test
Accuracy	0.7775	0.6667	0.7475	0.6667	0.7258	0.6852	0.7559 [#]	0.6667	0.9884	0.9844 [*]
Precision	0.5448	0.5485	0.5988	0.6265	0.6358 [#]	0.6285	0.5998	0.5445	0.7442	0.7321 [*]
Recall	0.5445	0.5885	0.4836	0.5668	0.6585 [#]	0.6142	0.5544	0.5322	0.6855	0.7855 [*]
Specificity	0.8988	0.8066	0.8995	0.6999	0.7369	0.8426	0.8965 [#]	0.6582	0.9456	0.9854 [*]
F-measure	0.8638	0.8175 [#]	0.8945	0.8758	0.8123	0.7486	0.7233	0.7421	0.8998	0.8748 [*]

Note: The maximum values are marked with *, and the second-highest values are marked with #.

Table 4.
Overall classification results in DSLR camera.

Perf. Measure	Decision Trees		<i>k</i> Nearest Neighbor						CNN	
	Train	Test	Train (k=3)	Test (k=3)	Train (k=4)	Test (k=4)	Train (k=5)	Test (k=5)	Train	Test
Accuracy	0.7111	0.6038	0.6985 [#]	0.6415	0.6984	0.6604	0.7001	0.6226	0.9645	0.9688 [*]
Precision	0.6642	0.6255	0.6185	0.6256	0.6358	0.6188	0.6123	0.6554 [#]	0.7589	0.7425 [*]
Recall	0.5645	0.5955	0.5036	0.5058	0.7155	0.6252 [#]	0.5444	0.5822	0.7555	0.7985 [*]
Specificity	0.7588	0.7566	0.7495	0.7399	0.7468	0.7585	0.7998 [#]	0.7214	0.9122	0.9258 [*]
F-measure	0.8466	0.8255	0.8756	0.8648 [#]	0.8087	0.7845	0.7158	0.7435	0.8798	0.8848 [*]

Note: The maximum values are marked with *, and the second-highest values are marked with #.

5. Discussion & Future Work

This research work was a preliminary investigation of Neonatal Hyperbilirubinemia detection using mobile devices, proposing a new mCADx to overcome the limitation of previous work. However, there are vital points that were discussed and addressed.

C. Efficient Image Data Preprocessing and Feature Selection

A similar work, BiliCam, performed a screening test on the subjects, particularly in infant cases. The limitation of the BiliCam is the application of a color calibration card and the BiliCam mobile application used to detect and monitor Neonatal Hyperbilirubinemia. BiliCam claimed an accuracy of 85% from 100 newborns. We investigated the robust image data

preprocessing techniques, including RGB, HSV, and LCH. Selected features were used in the classification model. In terms of classification results, our research work gains higher accuracy in all mobile devices and the DSLR camera without using the color calibration card. Specifically, CNN gained the best classification result at 0.9688, 0.9844, and 0.9688 in iOS, Android, and DSLR cameras, respectively. Two factors possibly affect this phenomenon. First, better resolution devices could improve the quality of images or the application of image data preprocessing and feature selection methods. Apart from reporting accuracy from the previous work, other performance measures must be considered in medical research problems. Our research work showed the most related performance measures, including precision, recall, specificity, and F-Measure [16].

Recently, a non-invasive screening device for Neonatal Hyperbilirubinemia was introduced. It analyzed the spectrum of an infant's nail bed for detecting jaundice severity level. The result was promising. However, this device is still in the early stage of development, and its application involved the use of special light beam generators, as opposed to our research work, which used only smart devices. Due to different environmental settings, comparing classification performances was challenging [17].

D. Evaluation and Comparison of Classification Models

BiliCam and our research work selected similar classification measurements to determine accuracy, however, we have added other necessary measurements for our proposed methods. Those measurements include Precision, Recall, Specificity, and F-Measure, though other research works may still employ other performance measurements that may be more suitable to their works. With the differences in performance measurements, it can be a challenge to compare research works. In addition, research works utilizing the data mining model have usually used the k-folds cross-validation technique for model evaluation.

E. Practical Use of the Application

Our research work intended to develop the mCADx installed on mobile phones or tablets to be used as a screening tool aligned with BiliCam's purpose. Other research works intended to be an advanced diagnostic tool used or to replace existing tools. With the purpose of this research work, balances between accuracy and processing time play important roles. In terms of F-Measure, the DT won in two cameras, iOS and Android, which was our attention. The DT showed a promising result as we plan to develop the mCADx. In addition, the DT processing time was reasonable.

F. Limitation and Future Work

The number of infant skin photo datasets was limited in our hospital. All collected photos are of Asian infants, mainly in the southern part of Thailand. An expansion of datasets to include other races or subjects in a different part of the world may improve our classification performance results. In terms of classification, multifractal methods and other hybrid data mining techniques can be considered for future study. Processing time is another dimension that can be used as one of the performance measures, which will be added to our future work.

References

- [1] B. O. Olusanya, F. B. Osibanjo, and T. M. Slusher, "Risk factors for severe neonatal hyperbilirubinemia in low and middle-income countries: A systematic review and meta-analysis," *PLoS One*, vol. 10, p. e0117229, 2015. Available at: <https://doi.org/10.1371/journal.pone.0117229>.
- [2] N. Ambalavanan and W. Carlo, "Nelson textbook of pediatrics, Jaundice and hyperbilirubinemia in the newborn, R. Kliegman, B. Stanton, J. III, N. Schor and R. Behrman, Eds," ed Philadelphia: Elsevier, 2016, pp. 871-879.
- [3] A. E. Burgos, V. J. Flaherman, and T. B. Newman, "Screening and follow-up for neonatal hyperbilirubinemia: A review," *Clinical Pediatrics*, vol. 51, pp. 7-16, 2012. Available at: <https://doi.org/10.1177/0009922811398964>.
- [4] P. Woodgate and L. A. Jardine, "Neonatal jaundice," *BMJ Clinical Evidence*, p. 0319, 2011.
- [5] M. Kaplan, R. Wong, J. Burgis, E. Sibley, and D. Stevenson, "Fanaroff and Martin's neonatal-perinatal medicine diseases of the fetus and infant, Neonatal jaundice and liver diseases, R. Martin, A. Fanaroff and M. Walsh, Eds," ed Philadelphia: Elsevier, 2020, pp. 1788-853.
- [6] J. F. Watchko, *Neonatal indirect hyperbilirubinemia and kernicterus. In Avery's Diseases of the Newborn*: Elsevier, 2018.
- [7] B. O. Olusanya, M. Kaplan, and T. W. Hansen, "Neonatal hyperbilirubinaemia: A global perspective," *The Lancet Child & Adolescent Health*, vol. 2, pp. 610-620, 2018. Available at: [https://doi.org/10.1016/s2352-4642\(18\)30139-1](https://doi.org/10.1016/s2352-4642(18)30139-1).
- [8] V. Bhutani, R. Vilms, and L. Hamerman-Johnson, "Universal bilirubin screening for severe neonatal hyperbilirubinemia," *Journal of Perinatology*, vol. 30, pp. S6-S15, 2010.
- [9] N. Bosschaart, J. H. Kok, A. M. Newsum, D. M. Ouweneel, R. Mentink, T. G. van Leeuwen, and M. C. Aalders, "Limitations and opportunities of transcutaneous bilirubin measurements," *Pediatrics*, vol. 129, pp. 689-694, 2012. Available at: <https://doi.org/10.1542/peds.2011-2586d>.
- [10] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, and C. Shearer, "CRISP-DM 1.0 Step-by-step data mining guide. Retrieved from <https://www.the-modeling-agency.com/crisp-dm.pdf>." 2020.
- [11] J. Han, M. Kamber, and P. Jian, *Data mining concepts and techniques*. San Francisco, CA: Morgan Kaufmann Publishers, 2011.
- [12] R. Ihaka, "Colour for presentation graphics," in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing, Vienna, Austria*, 2003.
- [13] A. Zeileis, K. Hornik, and P. Murrell, "Escaping RGBland: selecting colors for statistical graphics," *Computational Statistics & Data Analysis*, vol. 53, pp. 3259-3270, 2009. Available at: <https://doi.org/10.1016/j.csda.2008.11.033>.
- [14] R. Stauffer, G. J. Mayr, M. Dabernig, and A. Zeileis, "Somewhere over the rainbow: How to make effective use of colors in meteorological visualizations," *Bulletin of the American Meteorological Society*, vol. 96, pp. 203-216, 2015. Available at: <https://doi.org/10.1175/bams-d-13-00155.1>.

- [15] International Color Consortium, "Specification ICC.1:2004-10 (Profile version 4.2.0.0) image technology colour management — architecture, profile format, and data structure," 2006.
- [16] L. Greef, M. Goel, M. J. Seo, E. Larson, M. M. Stout, M. Taylor, and S. Patel, "Bilicam: Using mobile phones to monitor newborn jaundice," presented at the UbiComp '14: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 2014.
- [17] A. Halder, M. Banerjee, S. Singh, A. Adhikari, P. K. Sarkar, A. M. Bhattacharya, and S. K. Pal, "A novel whole spectrum-based non-invasive screening device for neonatal hyperbilirubinemia," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, pp. 2347-2353, 2019. Available at: <https://doi.org/10.1109/jbhi.2019.2892946>.