Artificial Intelligence In Medicine 162 (2025) 103088

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

Artificial Intelligence non-invasive methods for neonatal jaundice detection: A review☆

Fati Oiza Salami [a],*, Muhammad Muzammel [a], Youssef Mourchid [b], Alice Othmani [a]

[a] Laboratoire Images, Signaux et Systémes Intelligents (LiSSi) EA 3956, Université Paris Est Créteil (UPEC), 122 Rue Paul Armangot, Vitry Sur Seine, Créteil, 94010, France

[b] CESI LINEACT Laboratory, UR 7527, Dijon, 21000, France

ARTICLE INFO

Keywords:
Neonatal jaundice
Non-invasive detection
Machine learning
Deep learning
Bilirubin

ABSTRACT

Neonatal jaundice is a common and potentially fatal health condition in neonates, especially in low and middle income countries, where it contributes considerably to neonatal morbidity and death. Traditional diagnostic approaches, such as Total Serum Bilirubin (TSB) testing, are invasive and could lead to discomfort, infection risk, and diagnostic delays. As a result, there is a rising interest in non-invasive approaches for detecting jaundice early and accurately. An in-depth analysis of non-invasive techniques for detecting neonatal jaundice is presented by this review, exploring several AI-driven techniques, such as Machine Learning (ML) and Deep Learning (DL), which have demonstrated the ability to enhance diagnostic accuracy by evaluating complex patterns in neonatal skin color and other relevant features. It is identified that AI models incorporating variants of neural networks achieve an accuracy rate of over 90% in detecting jaundice when compared to traditional methods. Furthermore, satisfactory outcomes in field settings have been demonstrated by mobile-based applications that use smartphone cameras to estimate bilirubin levels, providing a practical alternative for resource-constrained areas. The potential impact of AI-based solutions on reducing neonatal morbidity and mortality is evaluated by this review, with a focus on real-world clinical challenges, highlighting the effectiveness and practicality of AI-based strategies as an assistive tool in revolutionizing neonatal care through early jaundice diagnosis, while also addressing the ethical and practical implications of integrating these technologies in clinical practice. Future research areas, such as the development of new imaging technologies and the incorporation of wearable sensors for real-time bilirubin monitoring, are recommended by the paper.

Contents

☆ This document is the outcome of the research study undertaken at Laboratoire Images, Signaux et Systémes Intelligents (LiSSi) EA 3956, Université Paris Est Créteil (UPEC), 94010, Créteil Cedex, France.

* Corresponding author.
E-mail addresses: fati.salami@u-pec.fr (F.O. Salami), muzammel88@gmail.com (M. Muzammel), ymourchid@cesi.fr (Y. Mourchid), alice.othmani@u-pec.fr (A. Othmani).

https://doi.org/10.1016/j.artmed.2025.103088
Received 17 September 2024; Received in revised form 27 January 2025; Accepted 11 February 2025
Available online 19 February 2025

1. Introduction

Jaundice is a prevalent indication of a medical condition that is defined by a yellowish coloring of the skin, the white part of the eyes (scelera), and other mucous membranes. It is caused by high amounts of bilirubin in the blood, also known as hyperbilirubinemia, which can result from various factors such as alcoholic liver disease, hepatitis, hemolysis, or bile duct obstruction. Jaundice can be life-threatening for both newborns and adults, and it is challenging to detect even for experienced physicians, with a detection rate of approximately 70 percent. [1,2]. In newborns, this condition is known as neonatal jaundice and it usually occur within the first week of life due to the neonate's immature liver, which struggles to process and eliminate bilirubin effectively [3].

Symptoms of excessive sleeping, leading to poor feeding and nutrition, are often exhibited by newborns with jaundice, affecting their overall well-being [4]. Neonatal jaundice is mostly responsible for hospitalization or readmission of newborns for intensive care in the first week of life, in the absence of timely diagnosis and treatment, fatality or permanent brain damage may occur [3]. Up to 60% of term babies and 80% of preterm babies globally were reported to have developed neonatal jaundice within their first week of life [5,6] and 10% of newborns that were breastfed developed jaundice up to a maximum of four weeks old. [7,8]. This common occurrence of jaundice in newborns can be attributed to physiological factors related to bilirubin metabolism.

Bilirubin is a yellowish pigment that is produced when there is a breakdown of old red blood cells, these red blood cells survive for a shorter time in newborns than they do in adults, which is about 120 days [9,10]. As a result, newborns have a significantly elevated amount of red blood cells that must be broken down, resulting to increased bilirubin levels [4]. Jaundice may become significantly visible when serum bilirubin level exceeds 2.0 to 2.5 mg/dl. [9] in their study states that in newborn infants, a bilirubin level over 5 mg/dL (85 umol/l) indicates clinical jaundice, whereas, jaundice is indicated in adults at a level of 2 mg/dL (34 umol/l). Neonates with jaundice exhibit a distinct icteric sclera, yellowing of the forehead and sternum. The onset of jaundice first becomes apparent on the newborn baby's face, then spreads downwards to the trunk and then to the limits as the bilirubin level rises. [9]. Short-term excess bilirubin is harmless, but a high level in newborns can lead to Kernicterus, a chronic form of bilirubin encephalopathy, which can cause cerebral palsy, deafness, language difficulties, developmental delay, or even death [4,7] as highlighted in Fig. 1.

Neonatal jaundice is divided into two categories; Physiological jaundice which becomes evident between 24 to 72 h after birth and disappears between 10 to 14 days of life [7] and pathologic jaundice which is the most severe type of jaundice, it appears within 24 h after birth, and is often associated with a quick rise in the bilirubin levels of newborns due to the elevated breakdown of red blood cells [7]. Given the differences in onset and severity between physiological and pathologic jaundice, accurate and timely diagnosis is crucial for effective management.

A combination of clinical, medical, and non-invasive methods is typically employed to diagnose neonatal jaundice. By visually observing the color of the skin and sclera of the eye, seasoned medical practitioners can clinically detect neonatal jaundice [11,12], which is characterized by yellowish discoloration of the skin and eyes when bilirubin levels exceed specific thresholds [13]. However, medical approaches involving invasive blood tests to ascertain the actual levels of Total Serum Bilirubin (TSB) has to be done to validate the observations recorded. A syringe is used to collect blood from a vein in the patient's arm, the sample is subsequently tested to determine the bilirubin level in the patient's blood. Typically, the patient will feel minor discomfort and a bruise where the needle was inserted [13]. This approach is considered to be the gold standard but it is invasive and causes pain and discomfort for the neonate [4,6]. This invasive technique causes discomfort for the infant, raises the possibility of infection at the sampling site, and worries the parents. However, the non-invasive approach is fast, effortless to use, and painless, which is particularly beneficial for neonates.

Non-invasive methods includes transcutaneous bilirubinometry (TcB) applied to eyes, facial, and chest skin images [14,15], optical spectroscopy, and reflectance spectroscopy applied to thumb nail images [8,16]. These approaches offer less discomfort, lower infection risk and early detection, however, there are associated limitations with these approaches particularly regarding accuracy across different skin tones and the need for regular calibration [8,16]. In order to tackle some of the aforementioned limitations, some non-invasive approaches adopts the use of image processing tools. For instance, the work of [17] extracted RGB, YCbCr, and HSV color space features to classify neonatal images, [3] conducted semantic mask detection and scelera segmentation using LabelMe and [9] achieved face landmark detection and ROI selection using Histogram of Oriented Gradients.

Early detection of neonates on the verge of severe hyperbilirubinemia and kernicterus is highly crucial to ensure effective treatment, especially for their immature central nervous system particularly to prevent the adverse outcomes usually associated with hyperbilirubinemia [4]. Jaundice may not be apparent at the time of hospital discharge as many newborns are discharged within 48 h after birth. The early release from the hospital after the birth of the newborn has saddled parents with the responsibility to identify neonatal jaundice and promptly seek medical attention [3]. Mothers are often the first to notice jaundice, especially outside of hospitals. Maternal education and empowerment are essential for timely jaundice detection, particularly in low- and middle-income countries as it is crucial for effective interventions to address severe neonatal jaundice burden [18]. Building

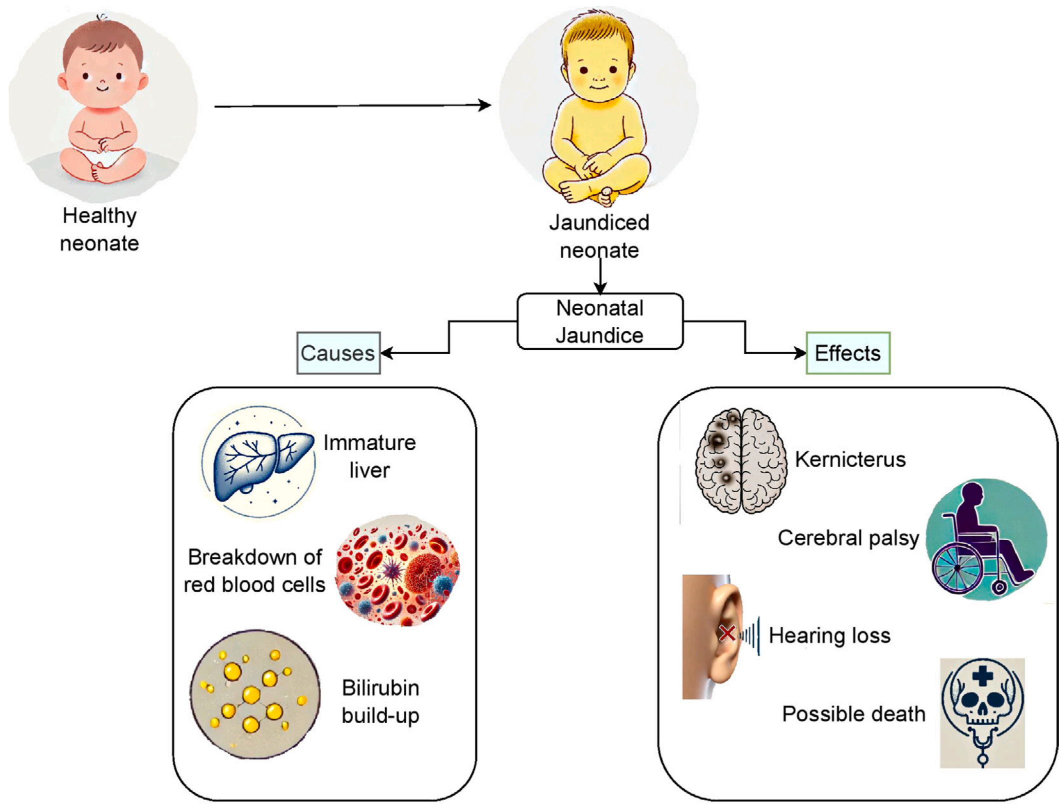

Fig. 1. The figure illustrates the main causes and effects of neonatal jaundice. The causes include an immature liver, breakdown of the red blood cells, and buildup of bilirubin, while the effects are kernicterus, celebral palsy, hearing loss, and death.

on the importance of early detection, several non-invasive methods have been developed to monitor bilirubin levels and manage neonatal jaundice effectively. Three broad categories of non-invasive approaches have been documented by previous research;

Transcutaneous Bilirubinometry (TcB) is used to estimate bilirubin levels in the blood by measuring the intensity of yellow pigmentation on the skin. A small device emits light on the forehead or sternum, bilirubin absorbs light in a specific spectrum, particularly in the blue and green ranges (around 450–500 nm). TcB is commonly used with devices like Bilicam [19] and Bilicapture [20]. However, potential inaccuracies can occur in certain populations, and the need to calibrate and validate the device's results against serum bilirubin levels makes it challenging to use. Optical spectroscopy measures bilirubin concentration by analyzing light interactions with the skin. It uses a device such as Bilicheck [21] to measure the intensity of light absorbed and scattered by skin components like melanin, hemoglobin, and bilirubin. Though it facilitates timely clinical decisions and is safe for newborns due to its low-intensity light, variations in skin pigmentation can affect accuracy. Reflectance spectroscopy measures the intensity of reflected light at different wavelengths by directing light at multiple wavelengths onto the skin and interacting with melanin, hemoglobin, and bilirubin [22]. The JM-103 is a commonly used reflectance spectroscopy device [14]. Although it eliminates the need for blood samples, variations in skin pigmentation can affect accuracy and require regular calibration.

While devices like the Bilicam, Bilicapture and JM-103 offer a non-invasive alternative to traditional blood tests, their reliance is limited by factors such as skin pigmentation, which can compromise accuracy. To address these challenges, recent research has increasingly turned to Artificial Intelligence (AI) to enhance the reliability of non-invasive methods. The focus of this research is on previous works that employed AI non-invasive methods for diagnosing neonatal jaundice, aimed at improved preventive care and treatment protocols for neonates. This research review is guided by the research questions given below:

1. What are the current Machine Learning (ML) and Deep Learning (DL) approaches adopted in Non-invasive Neonatal Jaundice Detection?
2. What is the extent of effectiveness of these identified Machine Learning and Deep Learning techniques deployed in Non-invasive Neonatal Jaundice Detection?
3. What are the identified shortcomings associated with using Machine Learning and Deep Learning techniques for non-invasive diagnosis of neonatal jaundice?
4. What performance evaluation metrics are utilized and how does it impact the performance assessment of the diagnostic tools?

1.1. Motivation and objectives

Jaundice in neonates is the prevalent reason of deaths and illness associated with newborns in the African continent and more specifically in West Africa. According to [6], 6 out of every 10 newborns are affected by jaundice. Hence, there is no overemphasizing the fact that jaundice is very deadly in neonates especially if not detected early enough. ML and DL approaches have demonstrated significant potentials for tackling this challenge [22–24]. Exploring artificial intelligence (AI) approaches for detecting neonatal jaundice is motivated by the prospects for an alternative non-invasive approach with enhanced accuracy and efficiency. AI algorithms can analyze complex patterns in data, distinguishing bilirubin levels from other variables such as skin pigmentation and lighting conditions [15,24].

Despite the substantial potential of AI, particularly deep learning models, to improve medical diagnosis and decision-making, its complete integration into clinical practice is impeded by a lack of readability and transparency [25,26]. It is essential to concentrate on methods that reduce uncertainty and clearly define AI predictions in order to increase acceptability. This ensures that AI supports human judgment rather than replaces it, particularly in complex fields such as medicine where skill, insights, and accountability are essential [27]. Personalized diagnostic results can be achieved by training these algorithms to consider for individual variations. Incorporating AI tools into diagnostic devices, enables the enhancement of their functionality, predict jaundice severity levels, and uncover new clinical insights. However, an objective assessment and analysis of the existing literature is crucial to provide insights into the most recent approaches, their effectiveness, constraints, and future research prospects.

An extensive overview of machine learning and deep learning-based applications for non-invasive neonatal jaundice detection is aimed to be offered to researchers, practitioners, and policymakers in this review. This will enhance their understanding of the field and guide towards the development of reliable and effective detection methods. Hence, this research makes several considerable contributions to the field of non-invasive neonatal jaundice detection as highlighted below;

1. In this work, we provided a comprehensive review of non-invasive technologies for diagnosing neonatal jaundice by assessing the traditional methods, such as the total serum bilirubin (TSB) measure in comparison to modern non-invasive approaches while highlighting the advancements and ongoing challenges in the field.
2. The study critically analyzed the application of machine learning (ML) and deep learning (DL) methodologies in diagnosing neonatal jaundice, identified their strengths and limitations, particularly in terms of accuracy, sensitivity, specificity, and clinical applicability.
3. This research work outlined key areas for future research, such as the development of novel imaging technologies like hyperspectral imaging, real-time bilirubin monitoring with wearable sensors, and the curation of diverse, annotated large-scale datasets.
4. It also addressed ethical and practical considerations for integrating these technologies into clinical practice, aiming to improve early detection and treatment of neonatal jaundice and reducing infant mortality related to this condition.

The following sections make up the rest of this paper: Section 2 details the review methodology, Section 3 discusses the data acquisition stage, Section 4 highlights the AI approaches and metrics, Section 5 covers the results and discussion of the study while Section 6 concludes the study.

2. Review methodology

In this review of non-invasive diagnosis of neonatal jaundice based on machine learning (ML) and deep learning (DL) techniques, the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guideline [28] was adopted, allowing for transparency and standardized documentation of the research method.

2.1. Search strategy

This study implemented a comprehensive search strategy across multiple online databases, notably IEEE Xplore, PubMed, ACM Digital Library, Scopus, Google Scholar, Science Direct, Academic Search, Web of Science and JSTOR, searching for relevant publications from January, 2014 to August, 2024. To identify these keywords, we relied on a set of predefined general items which are as follows: Neonatal Jaundice, Non-invasive Detection, AI approaches. A combination of keywords derived from the general items include: “non-invasive”, “jaundice diagnosis”, “neonatal jaundice” “neonates”, “jaundice detection”, “detection”, “machine learning” and “deep learning”. Table 1 presents the search phrases and keywords used. Development of the search strategy was achieved by consulting with the PRISMA guidelines [28]. A structured algorithm was utilized to carry out the search described in Table 2

2.2. Literature selection and inclusion criteria

The literature source selection process ensures transparency, reliability, and replicability in research. The research titles and abstracts of selected studies were carefully scrutinized by the reviewers to determine how suitable they are for this study. Fig. 2 illustrate the study selection process, following which the full text of prospectively eligible studies for inclusion in the review was obtained and a set of inclusion and exclusion criteria to identify suitable studies was developed. The inclusion covered studies conducted using machine learning, deep learning, and an ensemble of AI techniques for non-invasive neonatal jaundice detection, as well as studies that were peer-reviewed. Exclusions included studies focused on adult jaundice detection, diagnosing jaundice invasively, used non-AI techniques, studies with no abstract or full paper, published before January 2014, and not published in English. A total number of thirty-three (33) research articles satisfied the inclusion criteria as outlined in Fig. 3 by following the PRISMA guideline of [28]. The details of the articles considered for this review are shown in Table 3.

2.3. Information extraction

Information from the included studies has been extracted through a clear, reliable, and comprehensive process. This process systematically undertaken ensures that relevant pieces of information from the studies have been assessed for eligibility thereby enabling transparency and replicability in our review process. Information extracted include the study identification (authors, title, publication year, publisher), specific machine learning or deep learning techniques used, data type, main findings (results), performance metric adopted, challenges encountered, data management in terms of confidentiality and security of data etc. Table 4 provides a summary of the different methodological approaches adopted in the previous studies. The studies are divided into three categories; Machine learning-based techniques used in 17 studies, Deep learning-based techniques adopted in 7 studies and Ensemble methods and hybrid models used in 9 studies. Fig. 4 highlights the frequency of articles based on their publication year. It shows how many studies on non-invasive neonatal jaundice detection were published each year between 2014 and 2024 period been considered in this study. Fig. 5 illustrates the percentage distribution of studies by the category of technique used in previous non-invasive neonatal jaundice detection research works. The chart visually represents how many studies employed each category, highlighting the predominant use of Machine learning-based techniques over the other AI techniques in this research field.

3. Data acquisition

In the field of neonatal jaundice detection, understanding and correctly acquiring the right data is fundamental. In this section, the technologies employed in the process of acquisition of neonatal jaundice data is introduced, highlighting their advantages and limitations. In addition, emphasis was laid on the existing datasets that were obtained by using these technologies, and insights on the present state-of-the-art datasets are provided as a conclusion.

3.1. Technologies

In non-invasive neonatal jaundice detection research, several digital photography technology-based techniques have been deployed to collect image data. The tools as depicted in Fig. 6 include optical spectroscopic devices (e.g., SAMIRA) [16], smartphone applications (Billicam) [13], and reflectance photometers [15]. The technology adopted by these previous research typically focus on capturing images of the neonate’s skin, which can then be analyzed using AI-based techniques and methodologies to determine bilirubin levels and assess jaundice severity. The most common tools used by previous research have been highlighted.

Table 1
Approach for Keywords Search.

General Item	Keywords
Neonatal Jaundice	Non-invasive AND (Neonatal OR Neonates) AND (Detection OR Diagnosis OR Identification) AND (Jaundice OR Bilirubin)
Non-invasive Detection	(Neonate OR Neonatal) AND (Jaundice OR Bilirubin) AND (Diagnosis OR Identification)
AI approaches	(Machine Learning OR Deep Learning) AND (Neonatal OR Neonates) AND (Detection OR Diagnosis OR Identification)

Table 2
Algorithm for study search strategy.

Algorithm Search Strategy
1: **procedure** SearchStrategy 2: initialize electronic databases: IEEE Xplore, PubMed, ACM Digital Library, Scopus, Google Scholar, Science Direct, Academic Search and Web of Science 3: set search time frame: from January, 2014 to August, 2024 4: define keywords: “non-invasive”, “jaundice diagnosis”, “neonates”, “detection”, ”machine learning“ and ”deep learning" 5: adopt PRISMA guideline for conducting systematic reviews 6: implement database search using defined keywords 7: retrieve and categorize relevant publications 8: assess search results for inclusion based on review criteria (original research article, focused on non-invasive jaundice diagnosis, involving neonates, published in English, between January, 2014 to August, 2024. 9: **end procedure**

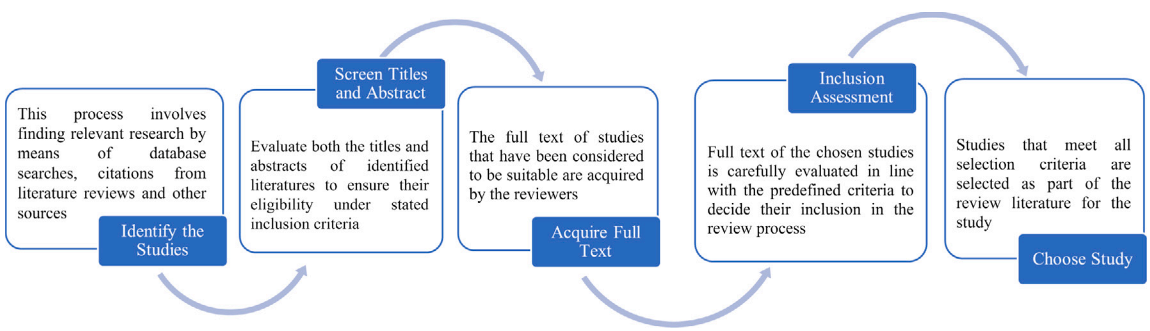

Fig. 2. Study selection process.

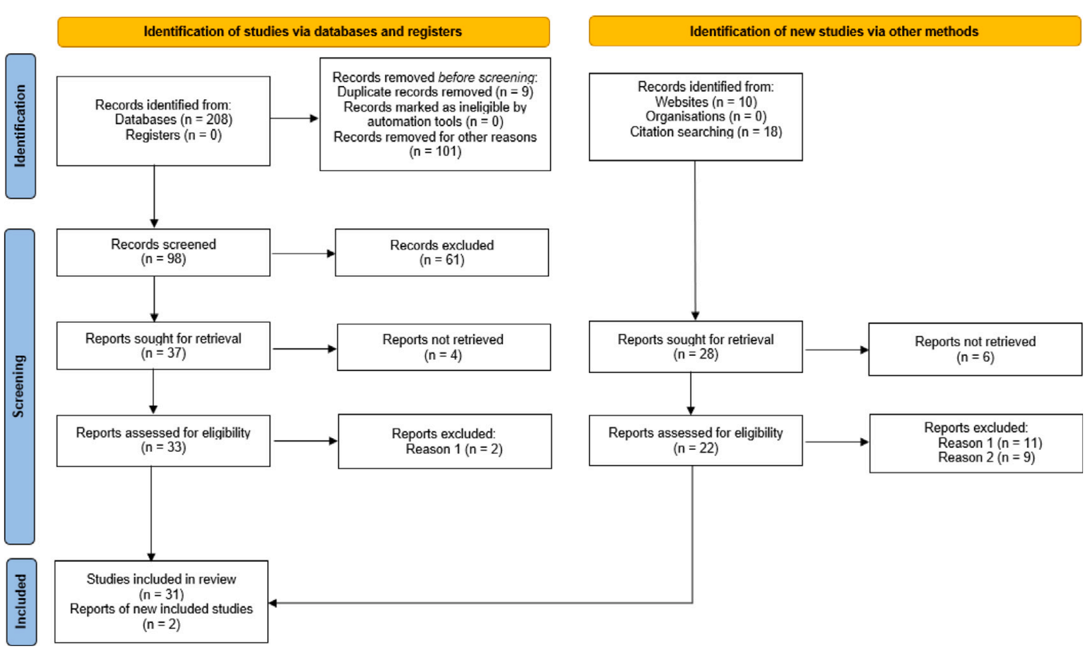

Fig. 3. Flow diagram adopted from the PRISMA guideline of [28] for the selection process of the 33 research articles based on the inclusion criteria.

Table 3
Details of selected studies organized chronologically, from the most recent to the least recent.

S/N	Study	Publication year	Publisher
1.	[29]	2024	Elsevier
2.	[30]	2024	Foundation of CS Inc.
3.	[3]	2023	Elsevier
4.	[9]	2023	Taylor and Francis
5.	[16]	2023	Springer
6.	[23]	2023	IEEE
7.	[31]	2023	MTU
8.	[32]	2022	Nova Science
9.	[33]	2022	Atlantis Press
10.	[34]	2022	Innovative Research
11.	[35]	2022	PLOS
12.	[10]	2021	IOP Publishing
13.	[17]	2021	MDPI
14.	[24]	2021	Springer
15.	[36]	2021	TUBITAK
16.	[37]	2021	MDPI
17.	[38]	2021	MDPI
18.	[39]	2021	BioMed Central Ltd.
19.	[40]	2021	IOP Publishing
20.	[8]	2020	Springer
21.	[14]	2020	IAEME Publication
22.	[22]	2020	Wiley–Blackwell
23.	[41]	2020	PLOS
24.	[4]	2019	Science
25.	[15]	2019	Optica
26.	[42]	2019	Springer
27.	[43]	2019	IEEE
28.	[44]	2019	TUMS
29.	[45]	2018	IEEE
30.	[46]	2018	IJCP
31.	[13]	2017	AAP
32.	[47]	2016	IJCSIS
33.	[48]	2016	Springer

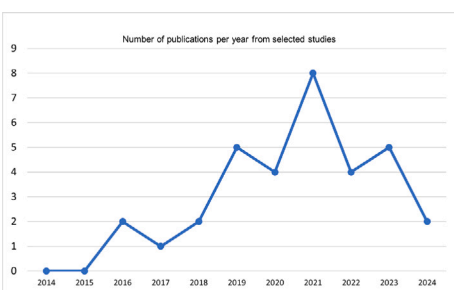

Fig. 4. Frequency of article based on publication year for selected studies.

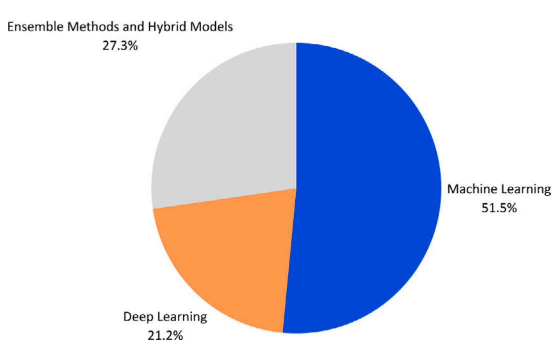

Fig. 5. Distribution of studies based on the category of AI technique used.

3.1.1. Digital cameras and smartphones

High-Resolution digital cameras are often used by researchers due to its affordability, ease of use and flexibility [15,17,37]. The digital cameras easily capture detailed images of the skin as these cameras are usually embedded with features like controlled lighting and macro lenses to enhance image quality. Modern smartphones equipped with high-resolution cameras are also often used for image capture [22,32,41]. They offer convenience and portability, making them suitable for use in clinical settings.

3.1.2. Hyper-spectral imaging systems

Hyper-spectral imaging (HSI) captures a broad spectrum of light, including wavelengths beyond the visible range, for each pixel in an image. This detailed spectral data allows for precise measurement of skin reflectance at different wavelengths [8]. HSI improves the accuracy of bilirubin estimation by identifying the distinct spectral signature of bilirubin, reducing interference from other factors like skin pigmentation and lighting [16]. By providing comprehensive spectral profiles, HSI enhances sensitivity to bilirubin and improves diagnostic accuracy. This technology offers a more reliable and non-invasive alternative to traditional methods for detecting neonatal jaundice.

3.1.3. Photometric devices

Photometric devices are usually smartphone-based application that uses the phone's camera in conjunction with a color calibration card to estimate bilirubin levels from images of the baby's skin. Examples of such devices is the Bilicam [19]. Table 5 presents the technology that has been adopted by the previous studies in acquiring data non-invasively for neonatal jaundice detection. As documented, 16 studies adopted the use of a smart phone camera with at least 8 MP and 1080 × 1920 resolution. 7 studies used a high definition digital camera ranging equipped with a macro lens, 2 studies used photometric devices named BilliCam and PiCam, 2 studies adopted Hyper-spectral imaging devices, 2 study used the MRI scanner and 1 study used the transducer to acquire ultrasound images of full-term babies.

3.2. Datasets

Non-invasive neonatal jaundice detection using AI techniques can be deployed using images from different parts of the body, including the scelera of the eyes, the face, chest, hands, palms, and even the thumbnail [49]. A few studies have successfully conducted non-invasive neonatal jaundice detection using clinical data; however, there is unfortunately limited literature focused on this data type as shown in Fig. 7, which highlights that image data has been predominantly used in previous research. This raises the question of whether it is more relevant to rely solely on image data, clinical health data, or a multimodal approach that integrates both to enhance diagnostic accuracy and robustness.

One major limitation of medical data is its limited availability and restricted access, hence, to the best of our knowledge, there is only one standard publicly available curated dataset for non-invasive jaundice diagnosis in neonates [29], this dataset consist of 767 neonates sternum skin images for both jaundiced and healthy babies. From the findings of this review, previous studies were conducted by sourcing for data from hospitals for the purpose of the research.

Due to the sensitive nature of these data, strict data management and confidentiality protocols were adopted, resulting in restricted access to research data in almost all the previous studies covered by this review. For instance, in studies that utilized eyes images such as [3,15,17,41] as well as studies that used face images, like those by [9,10,14,17,23,30,32,37,44,45,47]. Other dataset used in previous studies consisted of skin images as found in the studies of [13,22,30,31,33,34,36–38,40,46,48]. The study by [8,16,43] used thumbnail as inputs, [24,39] used MRI images in their study while [38] adopted the use of ultrasonic images.

Table 4
Categorization of studies based on techniques used.

Category	Number of Study	Technique Adopted
Category 1	17	Machine learning-based techniques
Category 2	7	Deep learning-based techniques
Category 3	9	Ensemble methods, Transfer learning and Hybrid models

Table 5
Technology deployed by the selected studies for data acquisition.

Data Capturing Technology	Studies	Cost and Accessibility
Smart Phone Camera	[3,13,17,22,29–34,40,41,44,46–48]	Affordable cost and easy accessibility.
HD Digital Camera	[9,10,29,34,37,40,45]	Relatively affordable and easy to access.
Photometric Device	[14,15]	Average cost and accessibility.
Hyper-spectral Imaging Device	[8,16]	High cost and not easily accessible.
MRI Scanner	[24,39]	Very high cost and low accessibility
Transducer	[35]	High cost and low accessibility.

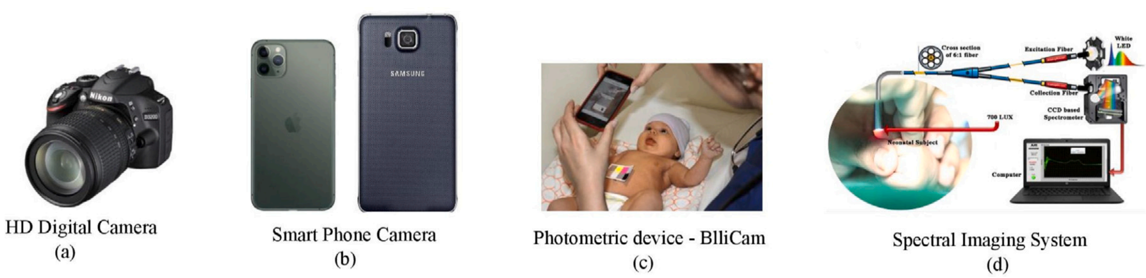

Fig. 6. (a) HD Digital Camera [15]; (b) Smart Phone Camera [31,48]; (c) Photometric Devices [13]; Spectral Imaging System [16].

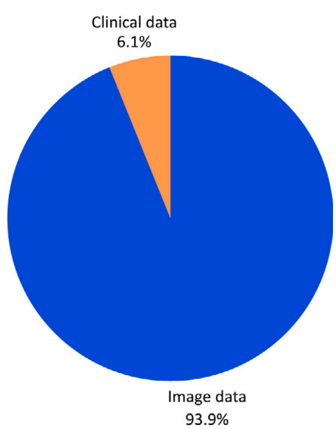

Fig. 7. Input data type used by previous studies detailing that 93.9% representing 31 studies used image data and 6.1% representing 2 studies that used clinical data.

3.3. Protocol for data collection

Based on the studies used in this review of non-invasive neonatal jaundice detection, the experimental setup involved the selection of the subjects for inclusion based on some criteria as well as collecting data from the selected subjects under specific and detailed conditions for data collection. As presented in Table 6, each study has collected their data based on specific subject inclusion criteria aimed at attaining the objectives of the study.

The Normal Jaundice Newborns (NJN) dataset, consisting of 767 infant skin images, was acquired at the NICU ward in Al-Elwiya Maternity Teaching Hospital in Baghdad, Iraq, as part of the study conducted by [29]. This consist of 567 healthy and 200 jaundiced images and the dataset is publicly available.

A face image dataset, consisting of 145 images (50 normal and 95 abnormal) of infants aged from one day to several weeks, was curated from Mosul hospitals (Alkhansaa Hospital and Ibn Alatheer Hospital) in the study by [32]. [17] collected a dataset consisting of eye images of a 100 neonates from King Khalid University Hospital (KKUH), Riyadh, Saudi Arabia. [33] used the chest skin images of 31 newborn infants with hyperbilirubinemia from the local district hospital. In the study of [22], chest skin images of 302 full-term, infants at up to 15 days of age were collected from 2 hospitals; St Olav Hospital, Trondheim and Akershus University Hospital, Lørenskog. [48] curated a chest images dataset of 40 jaundiced and 40 healthy newborns chosen from a patient group between 24 and 48 h after birth. [16] used thumb nail images from 3427 neonates with gestational age from 28 to 40 weeks for their study. [14] used an unspecified number of chest images of infants in their study.

In this study, [4] obtained vital health data covering 23 variables from 23 infants at a hospital in south-western Nigeria. [42] collected 2 datasets of neonatal data of neonates admitted to the University Children's Hospital Basel (UKBB) from 2015 to 2016 and records of neonates admitted to UKBB in 2017 within their first week of life and with a birth weight of at least 2500 grams. [34] curated chest skin image data from 178 infants from the newborn department at the Songklanagarind Hospital, Thailand. [43] collected thumbnail images as input data for their study from 1033 neonates from postnatal, neonatal intensive care unit and sick newborn care unit of Nil Ratan Sircar Medical College and Hospital, Kolkata. Also, [8] used thumbnail images of 1,968 neonates with gestational ages ranging from 28 to 40 weeks.

The study undertaken by [36] used skin images of 196 newborn babies, separated as control and case groups. [37] collected face and skin images from 20 infants, at the Ibn Al-Atheer teaching hospital of pediatrics in Mosul, Iraq. Ultrasound images of 180 patients from Taichung Veterans General Hospital was collected by [35] in their study while [38] acquired 228 chest images from 38 jaundiced neonates from

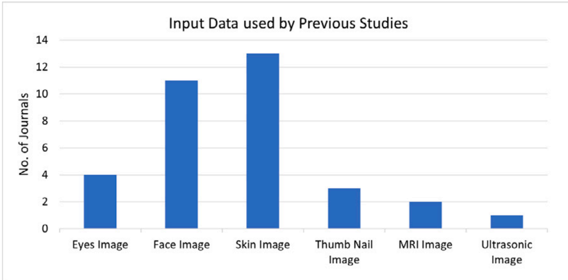

Fig. 8. Frequency of input data used by previous studies depicting that the skin image was the most frequently used input data by previous studies.

Chiayi Christian Hospital. A dataset of infant face images was obtained from University Hospital, University of Miyazaki by [45]. In the study by [9], 20 face images of 3D created babies with varying skin tones was used as input data. [15,41] recruited the face images of 87 newborn babies from the Neonatal Unit of UCL Hospital. [23] used face images of neonates from a number of children's hospitals in Chennai, Tamil Nadu, India.

Face images from 113 neonates who were at least 9 days old and weighed between 2.3 kg and 3.7 kg was acquired by [44]. [40] in their study, acquired thirty (30) samples of pictures of infant skin images. [47] used an unspecified number of face images in their research study. [3] annotated dataset of 201 eye images of 139 healthy neonates and 62 jaundiced neonates. [46] enrolled 35 newborn babies who were born at a hospital in Chennai and acquired their skin images. [13] obtained six BiliCam images each for 580 newborns with neonate age range of 12 to 163 h. [24] collected data from 150 male neonates, 75 with acute bilirubin encephalopathy (ABE) and 75 without ABE, who underwent MRI scans at post-menstrual age of 37–41 weeks. In the work of [39], 190 T1-weighted MRI image slices were collected from 79 neonates evenly distributed between ABE an d non-ABE cases. [31] acquired 511 images consisting of 100 images downloaded from the internet and 411 images acquired from Al-Elwiya Maternity Teaching Hospital, Al Rusafa, Baghdad. [30] in their study, used the NJN dataset for training the models while a dataset containing 19 faces images obtained from a GitHub repository was used for model inference testing. Fig. 8 highlights the frequency of input data as they were used in the previous studies.

Collecting data from the selected subjects in the previous studies was done under specific and detailed conditions. This is necessary in order to acquire high quality data needed for the analysis. Some studies observed specific criteria that are tailored to their research needs in their data acquisition process. For instance, [17,22,46] excluded subjects that were on phototherapy, had a birth weight less than 2 kg or had signs of diseases other than jaundice. [10,36,48] ensured the experiments were conducted in a controlled clinical environment to minimize external variables affecting the results. This includes controlled lighting and consistent ambient conditions to ensure accurate readings from the data. Images were captured with specialized devices such as HD camera and smartphones with high-quality cameras [15,33,34,44]. The devices are calibrated against standard references before use in the studies of [22,48]. In the studies of [17,31], detailed protocols for data collection were given, including the specific areas of the body where measurements are taken, the distance, and the angle of image capture to ensure consistency and reliability . Collected data were validated against traditional methods like serum bilirubin tests to assess the accuracy and reliability of the non-invasive methods.

The general process of data acquisition across the previous studies involved obtaining verbal and written consents from the parents of the neonates and obtaining ethical approval from the medical facility where the data acquisition takes place [13,22,46,48]. In this review, 10 studies in addition to the ethical approval, also adhered to the World Medical Association's Declaration of Helsinki guidelines for conducting research on human subjects [50]. Table 6 documents dataset features such as the size of dataset, the body part image used (eyes, face, chest skin, thumbnail images etc.).

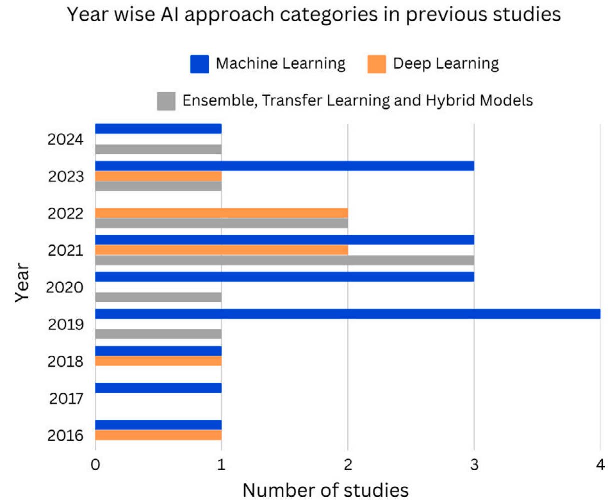

Fig. 9. Year-wise distribution of previous studies by the AI approaches used, highlighting that 2019 had the highest number of studies using ML, 2021 has the highest number of ensemble approach studies and 2022 had the highest number of studies conducted with deep learning.

4. Performance metrics and AI approaches

A wide range of techniques and evaluation criteria are used to develop, implement, and assess artificial intelligence systems. These approaches include machine learning algorithms and neural networks [3,9,23,32,45] while metrics includes accuracy, precision, recall, and F1-score [13,24,31,39]. Fig. 9 shows the distribution of previous studies by the AI approach adopted and their year of publication.

4.1. Performance measurement metrics

Previous studies used metrics such as accuracy, sensitivity, specificity, precision, and area under the receiver operating characteristic (ROC) curve (AUC). According to [17,24], these metrics are crucial in determining the reliability and clinical utility of non-invasive neonatal jaundice diagnostic tools and techniques, ensuring that they provide accurate, dependable, and timely identification of neonatal jaundice. By examining these performance measures, the study aims to provide a comprehensive assessment of current non-invasive techniques and their potential to replace or complement traditional invasive methods.

4.1.1. Accuracy

Accuracy refers to the overall correctness of a diagnostic method in identifying true cases of jaundice among newborns. It measures the proportion of true results (both true positives and true negatives) among the total number of cases examined [45]. [4] in their study, computed accuracy as shown in Eq. (1).

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} X\,100\% \tag{1}$$

Studies by [10,13] reported that high accuracy indicates that the diagnostic tool is effective in correctly identifying both jaundiced and non-jaundiced neonates, thereby minimizing misdiagnoses. Accuracy is a critical metric because it reflects the overall reliability and performance of the diagnostic method in clinical settings [36].

Table 6
Dataset used in previous studies detailing dataset references, description, modality, dataset accessibility status and the number of subjects.

Study	Dataset description	Modality	Public	Total no. of subjects
[3]	Annotated dataset of 201 eye images, 139 healthy neonates and 62 jaundiced neonates	Eye image	No	201
[4]	Vital health data covering 23 variables from 23 infants at a hospital in south-western Nigeria	Clinical data	No	23
[8]	1,968 neonates with gestational ages ranging from 28 to 40 weeks	Thumbnail image	No	1968
[9]	Created 20 images of 3D babies with varying skin tones	Face image	No	20
[10]	Images of 112 infants with at most 30 weeks gestational age from the Central Teaching Hospital of Pediatric, Baghdad, Iraq	Face image	No	12
[13]	Six BiliCam images were obtained each for 580 newborns with an age range of 12 to 163 h. Other features collected are ethnicity and race	Chest skin image	No	580
[14]	An image dataset with two class labeled as Jaundice and Non-Jaundice	Chest skin image	No	Not given
[15]	Data from 87 neonates were curated in the outpatient clinic of the Neonatal Unit of UCL Hospital	Eyes image	No	87
[16]	3689 neonates with gestational age from 28 to 40 weeks	Thumbnail image	No	3689
[17]	Dataset of a 100 neonates acquired from King Khalid University Hospital (KKUH), Riyadh, Saudi Arabia. Gestation age between 35 and 42 weeks, 62 male and 38 female newborns, with 67% healthy and 33% jaundiced.	Eye image and face image	No	100
[22]	302 full-term, healthy babies born at 37 weeks of gestational age or more and of normal weight of at least 2500 g and up to 15 days old from 2 hospitals	Chest skin image	No	302
[23]	Neonatal photos from a number of children's hospitals in Chennai, Tamil Nadu, India	Face image	No	Not given
[24]	Data collected from 150 HB neonates, 75 with ABE and 75 without ABE, who underwent MRI scans at postmenstrual age of 37–41 weeks, neonates were all male, weighed between 2.83 – 3.81 kg	MRI images	No	150
[29]	Normal Jaundice Newborns (NJN) dataset, captured at NICU ward in Al-Elwiya Maternity Teaching Hospital, Al Rusafa, Baghdad, Iraq. The dataset comprises of 767 infant images	Skin images	Yes	767
[30]	Normal Jaundice Newborns (NJN) dataset, captured at NICU ward in Al-Elwiya Maternity Teaching Hospital, Al Rusafa, Baghdad, Iraq comprising 767 infant images and a second dataset containing 19 face images obtained from a GitHub repository	Face and Chest skin image	Yes	779
[31]	511 images consisting of 100 images downloaded from the internet and 411 images acquired from Al-Elwiya Maternity Teaching Hospital, Al Rusafa, Baghdad	Chest skin image	No	511
[32]	Newborn image dataset of 145 images (50 normal and 95 abnormal) between the ages of one day and several weeks. Aggregated from Alkhansaa hospital and Ibn Alatheer hospital in Mosul	Face image	No	145
[33]	31 newborn infants with hyperbilirubinemia who were hospitalized in the local district hospital	Chest skin image	No	31
[34]	Data from 178 newborns from the Songklanagarind Hospital's newborn department, Thailand	Chest skin image	No	178
[35]	Ultrasound images of 180 patients from Taichung Veterans General Hospital	Ultrasound Images	No	180
[36]	196 newborn babies, separated as control and case groups	Skin image	No	196
[37]	Data set consisting of 20 infants, who are at least 48 h old and weighing 2500 g or more from Ibn Al-Atheer teaching hospital of pediatrics in Mosul, Iraq.	Face image and skin image	No	20
[38]	38 jaundiced neonates with 228 images at different color temperature from Chiayi Christian Hospital	Chest skin image	No	38
[39]	79 neonates diagnosed with hyperbilirubinemia (HB), including 47 with acute bilirubin encephalopathy (ABE) and 32 without ABE. A total of 190 slices (approximately 2–3 slices per patient)	MRI images	No	79
[40]	Thirty (30) samples of pictures of infant skin	Skin image	No	30

(*continued on next page*)

Table 6 (*continued*).

Study	Dataset description	Modality	Public	Total no. of subjects
[41]	Eighty-seven newborn babies from the Neonatal Unit of UCL Hospital	Eye image	Yes (Anonymized)	87
[42]	Dataset 1: Data of neonates admitted to the University Children's Hospital Basel (UKBB) throughout their first week of life in 2015 and 2016, respectively. Dataset 2: Records of neonates admitted to UKBB in 2017 in the first week of life and with gestational age more than 34 weeks	Clinical data	No	Not given
[43]	1033 newborns data from Nil Ratan Sircar Medical College and Hospital's postnatal, neonatal intensive care unit and sick newborn care unit, Kolkata	Thumbnail image	No	1033
[44]	113 neonate images, at least 9 days old, born less than 35 weeks' gestation and weighed between 2.3 kg and 3.7 kg	Face image	No	113
[45]	Dataset obtained from University Hospital, University of Miyazaki	Face image	No	Not given
[46]	35 neonates delivered at a hospital in Chennai were enrolled, neonates mostly term born, less than 1 week old, birth weight of 2500 grams or more from both male and female neonates	Skin image	No	35
[47]	Not stated	Face image	No	Not given
[48]	40 jaundiced and 40 healthy newborns chosen from a patient group between 24 and 48 h after birth	Chest skin image	No	80

4.1.2. Precision

Previous studies by [17,34] used precision to evaluate the performance of their techniques. It measures how many of the neonates identified as having jaundice actually have the condition. [31] reports that a high precision indicates that the diagnostic tool has a low rate of false positives, which means it is effective in correctly identifying jaundiced neonates without incorrectly labeling healthy ones as jaundiced. As shown in Eq. (2), [31] reported the following formula to compute precision.

$$\text{Precision} = \frac{TP}{TP + FP} X100\% \tag{2}$$

[17] emphasized in their study that precision is important for ensuring that treatment and further testing are only administered to those who truly need it, thereby reducing unnecessary interventions and associated costs.

4.1.3. Sensitivity

Sensitivity (also known as the true positive rate) refers to the ability of a diagnostic technique to correctly identify neonates who have jaundice. It measures the proportion of true positive cases (neonates with jaundice) out of the total actual positive cases (all neonates who truly have jaundice) [34]. High sensitivity indicates that the diagnostic tool is effective in detecting most of the cases of jaundice, hence, minimizing the risk of missing affected newborns. [44] in their study, computed accuracy as shown in Eq. (3) as:

$$\text{Sensitivity} = \frac{TP}{TP + FN} X100\% \tag{3}$$

Sensitivity is crucial for ensuring early and accurate identification of jaundice, as high sensitivity implies that even mild cases of jaundice are detected early which is important for timely intervention and treatment [23].

4.1.4. Specificity

Specificity (also known as the true negative rate) refers to the ability of a diagnostic method to correctly identify neonates who do not have jaundice [4]. As highlighted in Eq. (4), it measures the proportion of true negative cases (neonates without jaundice) out of the total actual negative cases (all neonates who truly do not have jaundice) [44].

$$\text{Specificity} = \frac{TN}{TN + FP} X100\% \tag{4}$$

According to [24], high specificity indicates that the diagnostic tool is effective in correctly identifying non-jaundiced newborns, minimizing the risk of false positives. Specificity is important for avoiding unnecessary treatment and anxiety for families of healthy newborns.

4.1.5. Area under the ROC curve (AUC)

The area under the receiver operating characteristic (ROC) curve, (AUC) is a performance measure that evaluates the overall diagnostic ability of a test. The ROC curve plots the true positive rate (sensitivity) against the false positive rate (1-specificity) at various threshold settings. The AUC represents the probability that the diagnostic method correctly distinguishes between a jaundiced and a non-jaundiced neonate.[23] The AUC can be computed using the trapezoidal rule, as demonstrated in Eq. (5).

$$\text{AUC} = \frac{1}{2}\sum_{i=1}^{n-1}(\text{FPR}(i) + \text{FPR}(i+1)) \cdot (\text{TPR}(i+1) - \text{TPR}(i)) \tag{5}$$

As reported by [24,42], a higher AUC value, closer to 1, indicates a better-performing diagnostic tool, with values close to 0.5 suggesting no diagnostic ability better than chance. An AUC of 1.0 denotes perfect accuracy, meaning the test perfectly discriminates between jaundiced and non-jaundiced newborns across all threshold levels [22]. The AUC is important for comparing different diagnostic methods and for assessing their potential effectiveness in clinical practice.

4.2. State-of-the-art in machine learning-based neonatal jaundice diagnosis

Machine learning-based neonatal jaundice diagnosis leverages advanced algorithms and data analytics to enhance the accuracy and efficiency of detecting jaundice in newborns. These cutting-edge techniques utilize a range of inputs, including clinical data and images, to provide timely diagnosis aimed at improved neonatal care. As depicted in Table 4, category 1 consisted of seventeen previous studies, [4,8,9,13,15,16,22,29,31,33,36,40,41,43,44,46,48] that adopted the use of varying machine learning techniques and tools in diagnosis of jaundice non-invasively.

A detailed overview of the studies in Category 1, including their research focus, methodologies, proposed solutions, and significance within the context of non-invasive neonatal jaundice detection, is presented in Table 7 of this research. These studies tackle several challenges including limited performance of existing jaundice detection systems, jaundice severity level diagnosis, intrusive methods of jaundice detection, lack of real time monitoring system and the need for improved strategies to detect neonatal jaundice. To overcome these challenges, solutions such as semantic masks detection, color map transformation, image segmentation techniques for color balancing, RGB color dispersion and threshold assessment, scleral chromaticity

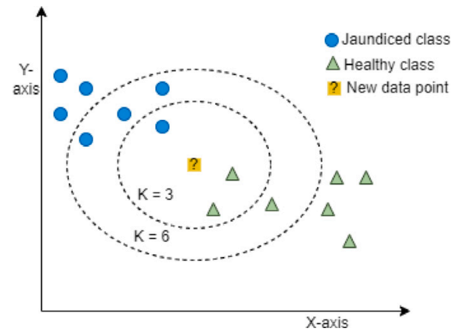

Fig. 10. K-Nearest Neighbor (KNN) model concept for classification.

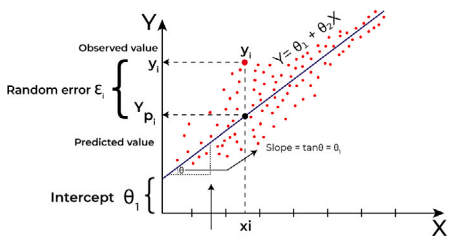

Fig. 11. Linear regression model concept for classification.

mapping to an SCB value were implemented using machine learning approaches such as random forest, support vector machine, K-nearest neighbor and linear regression among others. Fig. 10 [51] and Fig. 11 [52] illustrate the two most prevalent machine learning techniques; k-nearest neighbor (KNN) and linear regression used in the previous studies under review.

Findings from the studies highlights considerable advances in non-invasive diagnosis of neonatal jaundice. They demonstrate great accuracy, better detection rates, and out-performed some of state-of-the-art models. Furthermore, these works have present novel approaches for detecting the severity level of jaundice in neonates. Take for instance, [40] introduced the concept of Scleral Conjunctival Billrubin (SCB), in comparison with Transcutaneous Bilirubin (TcB) where the scleral chromaticity is mapped to an SCB value. [15] developed a novel color metric known as “Jaundice Eye Color Index” (JECI) that allows the yellowness of jaundiced sclerae to be quantified, [22] provided a billurubin level estimator from color calibrated smart phone images, [36] addressed the need for providing mobile support system to aid health care professionals remotely, an android-based efficient jaundice screening tool was introduced by [44] while [9] focused on skin color analysis using automatic ROI selection.

These research, which have proffered solutions to the challenges of existing techniques and established other novel approaches have given grounds for the availability of more efficient ways to diagnose neonatal jaundice non-invasively. For instance study conducted by [31] achieved an accuracy of 98.44%, [9] achieved an accuracy of 91.22%, [40] achieved an accuracy of 90.0% and [48] had accuracy of 85%. In this review, the most commonly used features are RGB values extracted form images of infant skin [31], Hue, Saturation, and Intensity (HSI) values derived from the RGB color space to better handle variations in lighting, Colormap transformations and feature calculation extracted from the important skin section [16,48], Region of Interest (ROI) [9], and estimated bilirubin levels. Other features used are age of neonates, birthweight and gestational age [4,44]. Studies conducted by [9,29, 48] adopted the use of multiple classifiers while other used single classifiers. [13] used the 10-fold cross-validation on the study sample.

4.3. State-of-the-art in deep learning-based neonatal jaundice diagnosis

Category 2 from Table 4 comprises of studies by [23,32,35,38,39, 45,47] that were conducted through the application of deep learning techniques in non-invasive diagnosis of neo-natal jaundice. The primary focus of these studies is to address the challenges associated with non-invasive diagnosis of jaundice in neonates, and they have proposed innovative deep learning-based solutions that leverage the strengths of network flow analysis, feature engineering, classifier training and optimization. These studies are especially relevant since they help to advance the state-of-the-art in non-invasive jaundice detection in neonates. The Deep Learning section in Table 8 presents insight into the comprehensive findings and methodologies adopted in the Category 2 studies, hereby underscoring their essential contributions to the field.

As highlighted in Fig. 12 which depicts the CNN model [53], a fore runner for deep learning models exploits the expansive amount of information that is the driving force behind network flow data, showcasing the extent to which it can encapsulate intricate patterns within data to achieve the aim of non-invasively diagnosing jaundice. As improved diagnostic performance ensures early jaundice detection in neonates.

In essence, the studies conducted in Category 2, highlights the importance of developing sophisticated techniques and tools for non-invasive jaundice detection. [38] for instance, proposed a dynamic threshold white balancing algorithm for effective jaundice diagnosis, [32] presented a sophisticated bilirubin monitoring system based on Artificial Neural Networks (ANN) while a non-contact and non-invasive approach for early detection was presented by [23,45] using spatial and spectral domain structures based on graph neural network (SSGNN), In [47], a system to categorize severity of jaundice by demarcation of the yellowness region of the skin using color segmentation was developed while [39] A stochastic gradient descent (SGD) optimizer with momentum was used to achieve optimized classification. With the results of studies in category 2 [23,35] showing a higher Area under the ROC Curve (AUC) than the category 1 studies, this implies that deep learning approaches outperform typical machine learning methods in terms of improved diagnostic accuracy and better reliability in differentiating between the classes being studied.

4.4. State-of-the-art in ensemble methods, transfer learning and hybrid models

Ensemble learning techniques and hybrid models are becoming a popular approach in enhancing the overall efficiency of jaundice non-invasive diagnosis systems, it involves the consolidation of varying individual models to create a synthesized model with greater precision and strength as a whole. Hybrid models on the other hand combine several integrate several machine learning algorithms, regularly incorporating conventional techniques with deep learning approaches, in order to exploit their complimentary capabilities.

Studies by [3,10,14,17,24,30,34,37,42] make up the category 3 found in Table 4, these studies have contributed significantly to the field of Non-invasive neonatal jaundice by being focused on overcoming challenges such as early jaundice detection, jaundice severity detection and identifying hyperbilirubinemia symptoms in neonates [17, 24]. The proposed solutions by the aforementioned studies include adopting a range of approaches such as transfer learning, Least Absolute Shrinkage and Selection Operator model, stacking and voting classifiers, gradient boosting and adaptive boosting [42]. Results from these studies display significant improvements in its application in early detection of neonatal jaundice non-invasively. The significance of these studies lies in their ability to improve the reliability and usability of the proposed techniques in addressing key issues such as the early detection of jaundice and the severity level estimation of jaundice in neonates [10,37]. In the Ensemble, Transfer Learning and Hybrid

Table 7
Key findings of machine learning-based studies detailing the study references, study approach, goal, and the performance metrics adopted by the studies.

Study	Approach	Goal	Performance Metrics
[4]	Deep learning classification and regression with multi-layer perceptron (MLP) classifier for varying number of epochs	Developed a classification model for predicting neonatal jaundice severity dependent on vital health records	60.9(Acc)
[8]	Positive predictive value, negative predictive value and constructing Receiver Operating Characteristic (ROC) curves	Higher accuracy in preterm infants and dark skin babies and low birth weight babies	–
[13]	Image collection, Bland–Altman analysis, mean bias and limits of agreement for the paired TcB and TSB measurements	Confirming BiliCam as an effective, user-adaptable app for the outpatient treatment and management of jaundiced infants	76.4(Se) 100(Sp) 85(AUC)
[15]	ROI selection, skin color analysis, extracting image features, and detection based on majority voting	A simple, non-invasive GUI system for tracking newborn jaundice	91.20(Acc)
[16]	Descriptive statistical analysis, simple linear regression analysis, and the Bland and Altman method	Developed a non-invasive point of care device — Spectrum Assisted Medical Inoffensive Radiation Application (SAMIRA)	95(Acc)
[22]	Generated a corresponding bilirubin estimate by comparing calibrated neonate images with a large database of bilirubin pairs	A neonatal jaundice screening tool that recorded high specificity in Caucasian newborn infants	100(Se) 69(Sp)
[29]	Skin detection, ROI selection, skin color analysis and training classifiers with intensity values	A non-invasive system adopting the use of a USB webcam and an XGBoost machine learning technique	99.39(Acc)
[31]	Skin color analysis, image histograms, used k-NN and SVR algorithms	A computer-assistive technologies to diagnose jaundice for different skin tones and light conditions	98.44(Acc) 97(Pr)
[33]	Image color intensity extraction using ImageJ software and correlated with hospital data for bilirubin concentration estimation	An easy to use method for neonatal hyperbilirubinemia estimation	–
[36]	Five-point body-part marker image analysis approach	An optical control system for detection of neonatal jaundice that can be deployed remotely	92.5(Acc)
[40]	Image complexion detection, RGB image processing and Euclidean distance measurement	Web-based digital image processing to detect jaundice cases based on bilirubin levels	90(Acc)
[41]	The scleral chromaticity mapping to a Scleral-Conjunctival Bilirubin (SCB) value	A novel screen-as-illumination ambient subtraction method to measure sclera chromaticity	100(Se) 42(Sp) 85(AUC)
[43]	Acquire and analyze optical signal and to calculate the bilirubin value	Non-invasive technique, successfully detects the bilirubin under various physiological conditions	88(Se) 98(Sp) 83(AUC)
[44]	Machine learning and regression-based techniques to analyze images pixel by pixel, estimate average RGB scores, convert images to HSI parameters	Smart phone-based estimation of bilirubin levels in infants	68(Se) 92.3(Sp)
[45]	Color spaces in native RGB and CIE XYZ colors and Jaundice Eye Color Index (JECI)	A predictive tool for the yellowness of the jaundiced scelera in relation to specific TSB levels	–
[46]	Biliscan application which utilized the smartphone's camera and a color calibration card	Mobile-based app with high correlation with the actual TSB values of the neonates adaptable to users	–
[48]	Pixel similarity and white balancing methods to obtain RGB and pixel values, feature extraction	A non-invasive jaundice diagnostic system based on SVR and KNN	85(Acc)

section of Table 8, a detailed overview of the studies classified as category 3 is presented, outlining their focus, methodologies adopted, and detailed findings. Transfer learning using a pre-trained VGC-16 model for feature extraction was deployed in [3], with a focus on input data mapping using non-linear features. [42] presented the LASSO model applied on Random Forest and Ensemble technique. The mCADx mobile computer aided diagnosis tool to identify neonatal hyperbilirubinemia was presented by [34,37] proposed a color transformation and Otsu thresholding technique to evaluate the B and Cb channels in neonatal images while [3] engaged in detecting the severity level of jaundice by augmenting images using cycleGAN.

An approach to training a deep convolutional neural network using images of unpaired dataset and Semantic Mask detection. [24] deployed multimodal MRI data apparent diffusion coefficient (ADC) maps combined with deep learning techniques for jaundice severity classification. In this study, a non-invasive method for detecting severe neonatal jaundice based on the NJN dataset was presented by [30], involving development of a system by improving the testing accuracy utilizing several variants of the YOLO object detection model. To improve the model's performance on the inference dataset (a dataset that is significantly different from the original dataset), the semantic segmentation image approach was implemented. The study compares the performance of custom CNN models and transfer learning models such as Vision Transformer, MobileNet, and EfficientNet across key metrics indicating that these techniques offer a cost-effective, accessible solution for early jaundice detection, especially in low-resource settings.

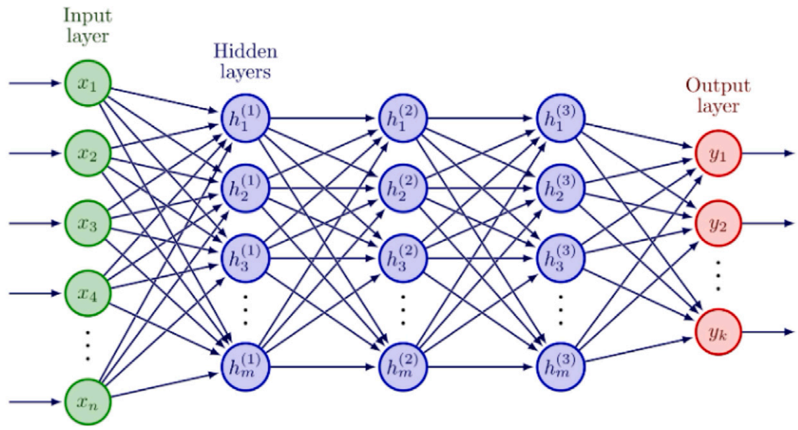

Fig. 12. CNN model concept for classification.

Table 8
Summary of Key Findings from Deep Learning-based Studies and Studies on Ensemble Methods, Transfer Learning, and Hybrid Models, highlighting Study References, Approaches, Goals, and Performance Metrics.

Study	Approach	Goal	Performance metrics
Deep Learning-Based Studies			
[9]	Skin area extraction, face area detection, converting RGB color values to YCbCr color space	Automatic detection system for neonatal jaundice by monitoring the changes in babies' skin color	–
[23]	Data pre-processing, segmentation, Processing with SSGNN	A novel framework based on graphical neural networks was developed	96.5(Acc) 96.6(Se) 95(Pr) 97.2(AUC)
[32]	Image augmentation, training on ResNet50	A Jaundice recognition system in newborn using deep learning	84.09(Acc)
[35]	Color temperature white balancing, image jaundice value detection	Reduce the impact of different color temperatures on images in the preprocessing stage	–
[38]	Use of ImageNet Large Scale Visual Recognition for classifying Ultrasound images as BA or non-BA	A screening tool for identifying possible BA by hepatobiliary ultrasound images	90.56(Acc) 67.8(Se) 96.9(Sp) 85.08(Pr) 92.6(AUC)
[39]	Use of ResNet18 fine-tuned on T1 weighted MRI dataset and optimized using stochastic gradient descent (SGD)	Advanced analytical methods to optimize T1 weighted images for diagnosing ABE in clinical practice	62.11(Acc) 78.95(Se) 45.26(Sp) 59.58(Pr) 68.92(AUC)
[47]	Smart phone-based image capturing, skin color segmentation	A non-invasive jaundice prediction system based on yellow discoloration of skin	93.0(Acc)
Ensemble, Transfer Learning and Hybrid Model Based Studies			
[3]	Image augmented using cycleGAN, ROI using Semantic Mask detection, color detection, scelera segmentation using LabelMe	Non-invasive early diagnosis of jaundice by analyzing eye images	88.57(Acc)
[10]	Skin image analysis using a color-based screening tool digital output transmission to a microcontroller circuit	Efficient and cost-effective skin color analysis-based jaundice detection and classification system	100(Acc)
[14]	Input data mapping using non-linear function. Color card bilirubin classification based on the skin and eye's coloration	Error-limited model to detect Neonatal Jaundice	93.0(Acc)
[17]	Classification and feature extraction by transfer learning with a pre-trained VGG-16 model	Transfer learning techniques performed better with skin attributes; standard machine learning models excelled with eye image features	86.83(Acc) 81.05(Se) 84.49(Pr) 82.12(F1) 81.05(AUC)
[24]	Multimodal MRI data, apparent diffusion coefficient (ADC) maps, combined with deep learning approaches to classify severity of jaundice(HB)	Demonstrated the potential of multimodal MRI with CNN models	93.0(Acc) 87(Se) 98(Sp) 98(Pr) 92.0(F1) 99.0(AUC)
[30]	Variants of the YOLO object detection model and semantic image segmentation on Transfer learning models including Vision transformer	Proposed approach has potential to serve as a cost effective alternative	100 (Acc) 100(Se) 100(Pr) 100(F1) 100(AUC)
[34]	Analysis of symptoms of neonatal hyperbilirubinemia utilizing image processing and data mining techniques	The use of sophisticated digital image processing and data mining techniques for detecting the symptom of neonatal hyperbilirubinemia	98.44(Acc) 79.85(Se) 98.54(Sp) 74.25(Pr)
[37]	Apply color transformation and Otsu thresholding methods to determine the values of the B channel and Cb channel for the ROI in the RGB and YCbCr color spaces	A highly efficient method for identifying jaundice at a TSB concentration of 14 mg/dL and above, with a detection time of 1 s	–
[42]	LASSO (Least Absolute Shrinkage and Selection Operator) model applied on random forest model, and an ensemble model	An early phototherapy prediction computational model	78(Se) 92(Sp) 95.20(AUC)

Table 9
Comparative overview for feature extraction in ML and DL approaches.

Study/Feature Extraction/ Representation	Strength	Limitation
Machine Learning		
Multi-space color feature extraction with colormap transformations [3,40,48]	Ability to capture diverse color characteristics relevant to the analysis	Increased computational load and time
Color channel mean extraction [9]	Detects jaundice with high sensitivity in mild cases	Lacks spatial information
RGB and YCrCb multi-color space models	Enhanced color analysis using intensity and luminance/chrominance [10]	Increased complexity due to multiple color models
Color-based feature representation [13]	Focused feature extraction for enhanced bilirubin estimation	Inaccuracies from calibration and skin tone variations
Reflectance photometry-based spectral data extraction [15]	High precision based on accurate spectral data	Increasing costs due to specialized instruments
Color calibrated images [22]	Provides a consistent and reliable image representation for analysis	Inconsistencies in the calibration process can introduce inaccuracies.
Lab color space transformation and OTSU thresholding [31]	Enhances color feature distinction, OTSU thresholding improves segmentation	Less reliable under different illumination conditions due to varying lighting
Deep Learning		
Graph-based spatial feature extraction using SPAGNN and SPEGNN [23]	Effectively captures spatial dependencies between data points	Computationally intensive, requiring significant resources
Multi-domain feature extraction technique [35]	Improve diagnostic by capturing intricate details	it may be computationally expensive
Color-space transformation and statistical feature extraction [47]	Effectively detects color changes and variability	Sensitive to lighting and skin tone variations

The solutions proffered in this category of research has highlighted the importance of developing robust non-invasive neonatal jaundice detection systems that can be assistive in early detection of jaundice and in estimating the severity level of jaundice in the neonates.

4.5. Comparison of ML and DL approaches

This section critically evaluates the efficacy and applicability of machine learning (ML) and deep learning (DL) methodologies in diagnosing neonatal jaundice non-invasively. This section explores how ML algorithms, which traditionally rely on feature engineering and statistical methods, compare with DL models that leverage complex neural networks for feature extraction and classification. It discusses the strengths and limitations of each approach concerning factors such as accuracy, interpretability, scalability, and data requirements. By examining these methodologies side by side, the study aims to provide insights into which approach may offer superior performance in clinical settings, thereby guiding future research directions and practical implementations in the field of neonatal care. Critically evaluates the efficacy and applicability of machine learning (ML) and deep learning (DL) methodologies in diagnosing neonatal jaundice non-invasively.

Feature Extraction and Representation. Feature extraction in machine learning is crucial for developing accurate diagnostic tools for neonatal jaundice. It involves transforming raw data into structured formats, identifying relevant clinical data, and using manually engineered features for input into models [33]. From reviewed studies, the type of features extracted for detecting neonatal jaundice includes color intensity of images [9] that help to detect jaundice even in very mild cases, RGB pixel values images [40,48] which enhances the early detection of jaundice, spectral data based on reflectance photometry [15], HSI parameters [44] that aid identification of neonates that require blood sample, color calibrated images [22] to identify severe jaundice with high intensity. In other studies, [31] achieved skin detection using the LAB color space transformation and the OTSU thresholding [10] for ROI selection, RGB and YCrCb color models was applied to the resulting skin image for skin color analysis. [3] conducted ROI segmentation with DeepLAB and color detection with delta E Distance metric in LAB space. To extract the RGB directly from the color filter array output data, [13] applied ROI and sensor Bayer pattern. Other image features that can be explored are alternative color spaces such as HSV and CMYK color spaces, biochemical markers like skin temperature, texture and structural features such as wavelet transform and Gabor filters.

Deep learning (DL) models, like CNNs and RNNs, learn hierarchical representations of input data, enhancing performance in non-invasive neonatal jaundice diagnosis by processing complex, non-linear relationships. In the deep learning approaches, [47] quantified yellowness of the facial skin by analyzing morphological changes in native RGB and CIE XYZ colors.

Study by [35] extracted hierarchical features from raw ultrasound images, focusing on both spatial and spectral domains. These feature extraction techniques help in capturing intricate patterns and features within images, improving the diagnostic accuracy. [23] extracted spatial and spectral domain facts from face and sclera images using the novel spatial and spectral graph neural network (SSGNN) model, that integrates spatial and spectral data from face and sclera images to enhance feature extraction. The use of SSGNN allows for a comprehensive analysis of both spatial and spectral information, capturing more detailed skin reflectance properties. Table 9 highlights the strengths and limitations of the feature extraction and representation approaches found in the machine learning and deep learning studies.

Model Complexity and Interpretability. Traditional ML techniques, which often rely on handcrafted features, can achieve good accuracy and precision but may struggle with complex patterns in the data because their performance depends on the quality and relevance of the features extracted manually. DL models, particularly Convolutional Neural Networks (CNNs), excel at extracting hierarchical features from raw data, capturing intricate patterns and details. This results in higher accuracy and precision in detecting neonatal jaundice, even in low severity cases. DL models can learn complex representations, making them more effective in differentiating between different severity levels of jaundice. ML models may require extensive pre-processing and feature engineering to balance the quality and variability of the input features, such as different lighting conditions, skin tones, and other variables; however, ML models generally require less computational power compared to DL models. They run efficiently on standard CPUs and do not require specialized hardware like GPUs. DL models are more robust to variations in input data due to their ability to learn from large, diverse datasets. They can automatically adjust to different lighting conditions, skin tones, and other factors, providing more accurate jaundice diagnosis; however, DL models require significant computational resources. Training these models necessitates powerful

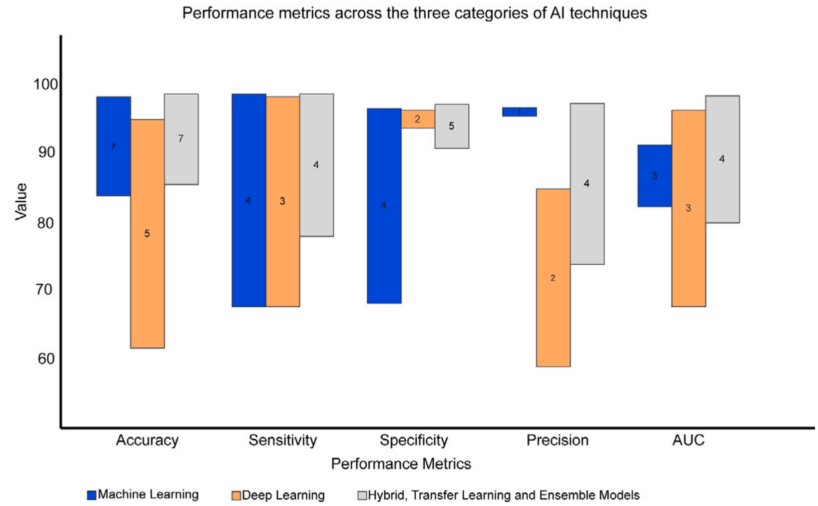

Fig. 13. Performance metrics used across the three AI techniques, including accuracy, sensitivity, specificity, precision and AUC. The top and bottom of the bars indicate the highest and lowest value achieved by each technique respectively.

GPUs or specialized hardware (like TPUs), as well as large amounts of memory and storage. In environments with limited access to resources, ML models are more suitable due to their lower computational and data requirements; they can be deployed on standard hardware and are more accessible in settings with constrained budgets. The high computational cost and the need for large datasets can be inhibiting factors in the adoption of DL models, especially in resource-constrained regions such as developing countries. The complexity of the models used in neonatal jaundice detection has a significant impact on both their performance and the resources required. Machine learning models offer a more accessible solution with lower computational and data demands but may not match the accuracy and robustness of deep learning models. Deep learning models, though they require substantial computational resources and large datasets, provide superior performance in early detection and higher accuracy in jaundice diagnosis by capturing intricate patterns in the data.

Performance Comparison. Comparing the performance metrics of machine learning (ML) and deep learning (DL) approaches is crucial for evaluating their effectiveness. Key metrics include accuracy, sensitivity, specificity, and area under the ROC curve (AUC). Accuracy of ML models is often high when suitable features are manually selected and engineered [33]. However, the performance may stagnate as these models heavily depend on the quality of input features. DL models tend to achieve higher accuracy by automatically learning complex patterns from raw data. This capability allows them to outperform ML models, especially when large datasets are available [23,32].

As presented in Fig. 13, ML models have good sensitivity [13], but their sensitivity depends on the features used. They may miss subtle jaundice patterns, resulting in lower sensitivity [44]. Deep learning models, on the other hand, have higher sensitivity due to their ability to capture complex relationships in data [23]. This is particularly useful in detecting jaundice where slight variations in skin color or texture can be critical [35]. ML models achieve high specificity with well-chosen features, but balancing sensitivity and specificity can be challenging [42]. Deep learning models improve specificity by effectively distinguishing between jaundiced and non-jaundiced neonates thereby reducing the likelihood of false positives, this is seen in studies by [23,35] with specificity of 96.9% and 95% respectively. ML models achieve high AUC with optimized feature selection and tuning, such as seen by [22] obtaining AUC of 92.5%and, but may not improve significantly with limited data capacity [41]. Deep learning models (DL) generally achieve higher AUC values, especially with larger datasets, indicating better diagnostic performance like in the studies of [23,35] achieving AUC of 92.9% and 97.2% respectively.

Transfer learning and ensemble models both showed significant improvements in the performance of neonatal jaundice detection. Transfer learning leveraged pretrained models to effectively extract relevant features, reducing the need for extensive data and computation specific to the task as seen in study of [24] with accuracy of 93%, specificity of 98% and AUC of 99.1%, as well as the study of [17] with an accuracy of 86.83% and AUC of 81.05%. Ensemble models enhanced robustness and accuracy by combining predictions from multiple models, ensuring reliable and consistent performance across different conditions such as in the studies of [42] with a specificity of 92% and AUC of 95.2% and [34] having accuracy of 98.44% and specificity of 98.54%. Both approaches demonstrated their utility in improving diagnostic accuracy and providing practical solutions for real-world applications.

The integration of diverse AI approaches and metrics demonstrates significant advancements in the field of neonatal jaundice diagnosis. Machine learning and deep learning techniques offer promising alternatives to traditional methods, with deep learning models showing superior performance due to their ability to handle complex patterns in raw data. The effectiveness of these AI systems is primarily assessed through key metrics such as accuracy, sensitivity, specificity, and AUC, which are crucial for ensuring reliable and timely diagnosis. Future research should focus on enhancing algorithmic robustness, developing comprehensive datasets, and exploring novel diagnostic modalities to further improve the efficacy and clinical adoption of these non-invasive technologies.

5. Discussion and challenges

This section offers a thorough analysis of the insights derived from our review, emphasizing on the methodologies employed for diagnosis of neonatal jaundice non-invasively through the utilization of ML and DL techniques. There are clear patterns which indicate to a preference for deep learning approaches, especially for deep neural networks, which are effective at extracting intricate details from unrefined data. ML techniques like decision trees, random forests, and support vector machines depicted high adaptability in handling heterogeneous feature spaces.

Predictive performance is improved by leveraging the strengths of individual ML or DL algorithms, which are fused through ensemble techniques [42]. Different types of algorithms or data modalities (such as combining image analysis with clinical data) are integrated in hybrid models to achieve more comprehensive and accurate diagnoses [14,24,34]. By amalgamating these approaches, there is promise in achieving higher precision in detecting neonatal jaundice, thereby

Table 10
Challenges in Machine Learning and Deep Learning non-invasive Diagnosis Techniques.

No.	Challenge	Description
1.	Accessibility to standard dataset	Previous research datasets are typically collected for specific purposes and cannot be shared with other researchers due to ethical agreements and privacy concerns of the subjects
2.	Skin pigmentation dependent	Previous studies that used skin image found their results less accurate in diagnosing jaundice in darker skin pigmentation.
3.	Class imbalance in dataset	This is common in medical datasets, in this case, the number of jaundiced subject will be far less than the number of healthy subject which would led to biased diagnosis
4.	Difficult data collection process	Data collection in hospitals, particularly from minors, is a tedious process requiring ethical approvals, consent forms, and potentially non-disclosure agreements, amidst bureaucratic processes
5.	Image Acquisition Environment	the illumination properties of the environment as well as the angle and distance of image capturing has an overall effect on the result of diagnosis
6.	Jaundice severity level dependent	In some studies, the technique applied has higher accuracy at specific correlated jaundice total serum bilirubin (TSB) measurement in the subjects

enhancing diagnostic reliability and clinical decision-making [37]. Despite significant advancements in technology, several factors impede the widespread adoption and effectiveness of non-invasive methods from previous research works. Primary concerns as highlighted in Table 10 includes the variability in skin pigmentation among neonates, the need for an extensive and diverse datasets to train and validate machine learning (ML) and deep learning (DL) models. Moreso, the integration of these diagnostic tools into clinical practice faces regulatory constraints, ethical considerations, and the necessity for real-time, reliable performance. Understanding these challenges is crucial for developing more robust, accurate, and clinically viable non-invasive diagnostic solutions for neonatal jaundice.

Dataset curated by most previous research were collected solely for the purpose of that research, and in most cases cannot not be shared with other researchers even when requested for [3,41]. This is due to ethical agreements reached and privacy concerns of the subjects [36]. In studies like [31,48] that used the skin image input, their techniques were more successful with lighter skin pigmentation [22,34]; hence, the techniques were not highly accurate in carrying out diagnosis on darker skin pigments. In [23], finding the yellowing signs in subjects with dark skin at the early onset of jaundice was quite difficult. In most medical datasets, class imbalance occurs in the curated dataset due to instances of one class (in most cases, the healthy class) significantly outnumbering the other class (typically the class with a medical condition or disease) [15,41]. This issue leads to poor performance, especially in machine learning techniques in predicting the minority class, and standard performance metrics may generate misleading results [3,34].

The process of data collection for research is a tedious task on its own, and it becomes even more challenging when it involves collecting data from human subjects within the confines of a hospital environment [22,29]. In this case data is collected form minors hence, the need to go through the hospital management's bureaucratic processes to obtain ethical approvals, obtaining consent forms and possibly signing non-disclosure agreements with parents. The environment in which image inputs are acquired plays a very significant role in the overall performance of the technique, though there are sophisticated image enhancing tools like in [34,48] that can be deployed to enhance the quality of the image. Poor illumination of the environment [10,37], variation in illumination, angle of capture, and distance of subject to the acquisition tool are some of the reoccurring challenges [14].

In some studies, the technique applied has higher accuracy at specific correlated jaundice total serum bilirubin(TSB) measurement which is obtained form the blood TSB test. [47] in their study developed a technique that detected jaundice only up to 20mg/dL while [13] could detect for TSB levels from 17 mg/dL and poor identification occurred for subjects with TSB levels higher than 15 mg/dL in the research of [16]. This implies that, these techniques will not be able to detect jaundice early on in the subject unless the hyperbilirubinemia builds up to a certain severity level or will only be able to detect jaundice when the hyperbilirubinemia level is very low.

The advancement and validation of non-invasive techniques for neonatal jaundice diagnosis have significant implications for healthcare policies and clinical practices. These studies underscore several key areas where changes and improvements can be implemented to enhance neonatal care: Adopting the use of AI-based tools for non-invasive bilirubin measurement in routine screening can help enhance early detection of neonatal jaundice, as well as enable frequent and painless monitoring of bilirubin levels in clinical practices. In this context, increasing AI acceptability in medical diagnosis requires addressing factors such as trust, system understanding, AI literacy, workflow integration, and socio-cultural concerns [54]. This highlights the need for human-centered AI systems tailored to healthcare professionals' expertise and norms, beyond just high algorithmic performance, to facilitate smoother data collection and integration into clinical workflows [26]. This will accelerate provision of continuous assessment and timely intervention without causing discomfort to the newborn. Non-invasive diagnostic tools empower early jaundice diagnosis, this can reduce reliance on serum bilirubin tests, lead to prompt treatment, minimizing infection risks and reducing hospital stays for affected infants and invariably reducing healthcare costs and resource burden. Furthermore, Advances in non-invasive technology pave the way for home monitoring of bilirubin levels, empowering parents to participate actively in their child's care and potentially reducing the need for returning to the hospital frequently. In health care policies, non-invasive jaundice screening and monitoring can be integrated into neonatal care guidelines for a standardized approach. Policymakers can create protocols mandating the use of validated devices. Healthcare professionals should be trained in using these tools, and educational programs can ensure accurate use across healthcare settings. These tools are especially beneficial in resource-limited settings, improving jaundice diagnosis and management, promoting healthcare access equity.

Summarily, by leveraging machine learning and deep learning techniques, non-invasive diagnosis of jaundice can present significant advantages;

1. We can analyze complex patterns in data that may not be apparent to humans, leading to more accurate diagnoses.
2. Enhance diagnostic precision by effectively differentiating between bilirubin levels and other variables, including skin pigmentation and illumination conditions.
3. Train Machine learning and Deep learning algorithms to account for individual variations, such as different skin tones and ages, providing personalized diagnostic results.

4. Improve the functionality of diagnostic devices such as improving the performance of transcutaneous bilirubinometers by integrating Machine learning and Deep learning algorithms into diagnostic devices.
5. Utilize machine learning and deep learning techniques for the processing of large volumes of data, which is essential for population-wide screening programs and epidemiological studies.
6. We can leverage Machine learning and Deep learning to predict the risk of developing severe jaundice based on early indicators, enabling proactive management and preventive care.
7. Adoption of Machine learning and Deep learning to uncover new insights from clinical data, contributing to the development of new diagnostic criteria and treatment protocols.

Policies and practices are significantly impacted by studies on non-invasive neonatal jaundice diagnosis, promoting an efficient, cost-effective, and compassionate approach to managing neonatal jaundice, with benefits for infants, parents, and healthcare systems. From the comprehensive finding in our study, research needs in the field can be in the following directions:

Improvement of model algorithms by developing advanced algorithms for better feature extraction from clinical data, ensuring more accurate representation of jaundice indicators. Directions can also come from the creation of robust models by exploiting further the strengths of techniques such as transfer learning and domain representation [17] that can performs efficiently across data acquired from different ethnic background and under varying clinical conditions [23,31]

Future research focus can be directed towards the use of novel imaging technologies, such as hyperspectral imaging [8,16], to capture more detailed and accurate information about skin pigmentation and bilirubin levels. Real time monitoring of bilirubin levels by development and validation of wearable sensors to enhance real time decision making can also be explored [9].

Curating standard, diverse and annotated large scale datasets that reflect various ethnicity, race and clinical scenarios to be used in training and validating ML and DL models effectively. Data is the most essential component in developing an efficient and robust ML or DL diagnosis tool. Based on previous research, only one standard dataset has been made publicly available for researchers in the field [17,23,29]. Research can be further promoted in the area of data sharing between institutions to facilitate collaborative research and accelerate the development of more accurate and generalizable diagnostic tools [3,34]. By addressing these areas, future research can overcome current limitations and enhance the efficacy, reliability, and adoption of non-invasive diagnostic technologies for neonatal jaundice.

6. Conclusion

In the study, the current landscape of non-invasive diagnostic technologies has been comprehensively examined, with significant advancements and ongoing challenges in this field highlighted. While traditional invasive methods, using the total serum bilirubin (TSB) measure, remain the gold standard [47] non-invasive approaches offer promising alternatives that can enhance early detection and improve neonatal care. Through critical analysis of both machine learning (ML) and deep learning (DL) methodologies, their respective strengths and limitations have been identified particularly in terms of accuracy, sensitivity, specificity, and real-world clinical applicability. For instance, detecting jaundice in neonates with high concentration in terms of Total serum Bilirubin (TSB) have been a major challenge especially for ML-based approaches. The review underscores the necessity for robust, diverse datasets and emphasizes the potential of ensemble techniques and hybrid models to further enhance diagnostic precision. Moreover, the discussion on regulatory and ethical considerations, alongside the practical challenges of clinical integration, points to the need for interdisciplinary collaboration and continued innovation. Future research should focus on improving algorithmic robustness, exploring new diagnostic modalities, and fostering collaborative efforts to develop reliable, scalable, and universally accessible non-invasive diagnostic tools for neonatal jaundice.

A number of limitations that could have an impact on the findings are acknowledged in this study; The insight acquired about the successful application of non-invasive diagnostic techniques may not be generalizable due to variations in the quality and sources of data used in various studies. Making direct comparisons between different ML and DL models used in different research is quite difficult due to variations in the datasets, assessment criteria, and implementation specifics. The rapid advancement in technology implies that certain approaches explored might soon become obsolete, while more recent ones might not be extensively explored. The review may not fully address all regulatory and ethical issues related to the deployment of non-invasive diagnostic tools, as these aspects often require specific analysis that are tailored to specific settings.

CRediT authorship contribution statement

Fati Oiza Salami: Writing – original draft, Methodology, Conceptualization. **Muhammad Muzammel:** Writing – review & editing, Supervision, Conceptualization. **Youssef Mourchid:** Writing – review & editing, Supervision, Conceptualization. **Alice Othmani:** Writing – review & editing, Supervision, Project administration, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We sincerely appreciate the financial support from the Nigerian Government, provided through the Tertiary Education Trust Fund (TETFund) 2023/2024 overseas scholarship award.

References

[1] Kihara T, Sugihara T, Ikeda S, Matsuki Y, Koda H, Onoyama T, et al. Identification and quantification of jaundice by trans-conjunctiva optical imaging using a human brain-like algorithm: a cross-sectional study. Diagnostics 2023;13(10):1767. http://dx.doi.org/10.3390/diagnostics13101767.

[2] Miah MMM, Tazim RJ, Johora FT, Al Imran MI, Surma SS, Islam F, et al. Non-invasive bilirubin level quantification and jaundice detection by sclera image processing. In: 2019 IEEE global humanitarian technology conference. Seattle, WA, USA: IEEE; 2019, p. 1–7. http://dx.doi.org/10.1109/GHTC46095.2019.9033059.

[3] Sreedha B, Nair PR, Maity R. Non-invasive early diagnosis of jaundice with computer vision. Procedia Comput Sci 2023;218:1321–34. http://dx.doi.org/10.1016/j.procs.2023.01.111.

[4] Chidozie EN, Onyenonachi Asinobi A, Adewunmi O, Aderounmu T, Ademola Adegoke S, Adebayo Idowu P. A classification model for severity of neonatal jaundice using deep learning. Am J Pediatr 2019;5(3):159. http://dx.doi.org/10.11648/j.ajp.20190503.24.

[5] Rennie J, Burman-Roy S, Murphy MS. Neonatal jaundice: Summary of NICE guidance. BMJ 2010;340(may19 3). http://dx.doi.org/10.1136/bmj.c2409, c2409–c2409.

[6] Olabisi A, Olawade D, Afolalu T, Wada O, Alabi D. Prevalence of jaundice among neonates admitted in a tertiary hospital in Southwestern Nigeria. Adv Pediatr Neonatal Care 2021. http://dx.doi.org/10.29011/APNC-121.100021.

[7] Dzulkifli FA, Mashor MY, Khalid K. Methods for determining bilirubin level in neonatal jaundice screening and monitoring: A literature review. J Eng Res Educ 2018;10:1–10.

[8] Halder A, Adhikari A, Ghosh R, Singh S, Banerjee A, Ghosh N, et al. Large scale validation of a new non-invasive and non-contact bilirubinometer in neonates with risk factors. Sci Rep 2020;10(1):11149. http://dx.doi.org/10.1038/s41598-020-67981-9.

[9] Khanam F-T-Z, Al-Naji A, Perera AG, Wang D, Chahl J. Non-invasive and non-contact automatic jaundice detection of infants based on random forest. Comput Methods Biomech Biomed Eng: Imaging Vis 2023;11(6):2516–29. http://dx.doi.org/10.1080/21681163.2023.2244601.

[10] Hashim W, Al-Naji A, Al-Rayahi IA, Oudah M. Computer vision for jaundice detection in neonates using graphic user interface. IOP Conf Ser: Mater Sci Eng 2021;1105(1):012076. http://dx.doi.org/10.1088/1757-899X/1105/1/012076.

[11] Riskin A, Tamir A, Kugelman A, Hemo M, Bader D. Is visual assessment of jaundice reliable as a screening tool to detect significant neonatal hyperbilirubinemia? J Pediatr 2008;152(6):782–7.e2. http://dx.doi.org/10.1016/j.jpeds.2007.11.003.

[12] Dionis I, Chillo O, Bwire GM, Ulomi C, Kilonzi M, Balandya E. Reliability of visual assessment of neonatal jaundice among neonates of black descent: A cross-sectional study from tanzania. BMC Pediatr 2021;21(1):383. http://dx.doi.org/10.1186/s12887-021-02859-x.

[13] Taylor JA, Stout JW, De Greef L, Goel M, Patel S, Chung EK, et al. Use of a smartphone app to assess neonatal jaundice. Pediatrics 2017;140(3):e20170312. http://dx.doi.org/10.1542/peds.2017-0312.

[14] Chakraborty A, Goud S, Shetty V, Bhattacharyya B. Neonatal jaundice detection system using CNN algorithm and image processing. Int J Electr Eng Technol. (IJEET) 2020;11(3). 348–264.

[15] Leung TS, Outlaw F, MacDonald LW, Meek J. Jaundice eye color index (JECI): Quantifying the yellowness of the sclera in jaundiced neonates with digital photography. Biomed Opt Express 2019;10(3):1250. http://dx.doi.org/10.1364/BOE.10.001250.

[16] Banerjee A, Bhattacharyya N, Ghosh R, Singh S, Adhikari A, Mondal S, et al. Non-invasive estimation of hemoglobin, bilirubin and oxygen saturation of neonates simultaneously using whole optical spectrum analysis at point of care. Sci Rep 2023;13(1):2370. http://dx.doi.org/10.1038/s41598-023-29041-w.

[17] Althnian A, Almanea N, Aloboud N. Neonatal jaundice diagnosis using a smartphone camera based on eye, skin, and fused features with transfer learning. Sensors 2021;21(21):7038. http://dx.doi.org/10.3390/s21217038.

[18] Olusanya BO, Slusher TM, Imosemi DO, Emokpae AA. Maternal detection of neonatal jaundice during birth hospitalization using a novel two-color icterometer. In: Szyld E, editor. PLoS One 2017;12(8):e0183882. http://dx.doi.org/10.1371/journal.pone.0183882.

[19] De Greef L, Goel M, Seo MJ, Larson EC, Stout JW, Taylor JA, et al. Bilicam: Using mobile phones to monitor newborn jaundice. In: Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing. Seattle Washington: ACM; 2014, p. 331–42. http://dx.doi.org/10.1145/2632048.2632076.

[20] Rizvi MR, Alaskar FM, Albaradie RS, Rizvi NF, Al-Abdulwahab K. A novel non-invasive technique of measuring bilirubin levels using BiliCapture. Oman Med J 2019;34(1):26–33. http://dx.doi.org/10.5001/omj.2019.05.

[21] De Luca D, Romagnoli C, Tiberi E, Zuppa AA, Zecca E. Skin bilirubin nomogram for the first 96 h of life in a European normal healthy newborn population, obtained with multiwavelength transcutaneous bilirubinometry. Acta Paediatr 2008;97(2):146–50. http://dx.doi.org/10.1111/j.1651-2227.2007.00622.x.

[22] Aune A, Vartdal G, Bergseng H, Randeberg LL, Darj E. Bilirubin estimates from smartphone images of newborn infants' skin correlated highly to serum bilirubin levels. Acta Paediatr 2020;109(12):2532–8. http://dx.doi.org/10.1111/apa.15287.

[23] Madhusundar N, Surendran R. Neonatal jaundice identification over the face and sclera using graph neural networks. In: 2023 5th international conference on smart systems and inventive technology. Tirunelveli, India: IEEE; 2023, p. 1243–9. http://dx.doi.org/10.1109/ICSSIT55814.2023.10060877.

[24] Wu M, Shen X, Lai C, You Y, Zhao Z, Wu D. Detecting acute bilirubin encephalopathy in neonates based on multimodal MRI with deep learning. Pediatr Res 2022;91(5):1168–75. http://dx.doi.org/10.1038/s41390-021-01560-0.

[25] Göndöcs D, Dörfler V. AI in medical diagnosis: AI prediction & human judgment. Artif Intell Med 2024;149:102769. http://dx.doi.org/10.1016/j.artmed.2024.102769.

[26] Catalina QM, Fuster-Casanovas A, Vidal-Alaball J, Escalé-Besa A, Marin-Gomez FX, Femenia J, et al. Knowledge and perception of primary care healthcare professionals on the use of artificial intelligence as a health-care tool. Digit Health 2023;9:20552076231180511. http://dx.doi.org/10.1177/20552076231180511.

[27] Lambert B, Forbes F, Doyle S, Dehaene H, Dojat M. Trustworthy clinical AI solutions: A unified review of uncertainty quantification in deep learning models for medical image analysis. Artif Intell Med 2024;150:102830. http://dx.doi.org/10.1016/j.artmed.2024.102830.

[28] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. PLoS Med 2021;18(3):e1003583. http://dx.doi.org/10.1371/journal.pmed.1003583.

[29] Abdulrazzak AY, Mohammed SL, Al-Naji A, Chahl J. Real-time jaundice detection in neonates based on machine learning models. BioMedInformatics 2024;4:623–37. http://dx.doi.org/10.3390/biomedinformatics4010034.

[30] Gupta K, Sharma V, Kathait SS. Smart screening: non-invasive detection of severe neonatal jaundice using computer vision and deep learning. Int J Comput Appl 2024;186.

[31] Yaseen Abdulrazzak A, Latif Mohammed S, Al-Naji A, Chahl J. Computer-aid system for automated jaundice detection. J Tech 2023;5(1):8–15. http://dx.doi.org/10.51173/jt.v5i1.1128.

[32] Abdulrhman AI, Jarjees MS, Nazar AK. Neonatal jaundice detection system using convolutional neural network algorithm. NeuroQuantology 2022;20(8):7536–46. http://dx.doi.org/10.14704/nq.2022.20.8.NQ44777.

[33] Anggraeni MD, Fatoni A, Rahmawati E, Nartiningsih I. Estimation of neonatal jaundice from the chest images captured with a smartphone:. In: Soedirman international conference on mathematics and applied sciences (SICOMAS 2021), vol. 5, Purwokerto, Indonesia: ATLANTIS Press; 2022, p. 23–6. http://dx.doi.org/10.2991/apr.k.220503.005.

[34] Dissaneevate S, Wongsirichot T, Siriwat P, Jintanapanya N, Boonyakarn U, Janjindamai W, et al. A mobile computer-aided diagnosis of neonatal hyperbilirubinemia using digital image processing and machine learning techniques. Int J Innov Res Sci Stud 2022;5(1):10–7. http://dx.doi.org/10.53894/ijirss.v5i1.334.

[35] Hsu F-R, Dai S-T, Chou C-M, Huang S-Y. The application of artificial intelligence to support biliary atresia screening by ultrasound images: A study based on deep learning models. PLoS One 2022. http://dx.doi.org/10.1371/journal.pone.0276278.

[36] Hardalaç F, Aydin M, Kutbay U, Ayturan K, Akyel A, Çağlar A, et al. Classification of neonatal jaundice in mobile application with noninvasive image processing methods. Turk J Electr Eng Comput Sci 2021;29(4):2116–26. http://dx.doi.org/10.3906/elk-2008-76.

[37] Hashim W, Al-Naji A, Al-Rayahi IA, Alkhaled M, Chahl J. Neonatal jaundice detection using a computer vision system. Designs 2021;5(4):63. http://dx.doi.org/10.3390/designs5040063.

[38] Hsu W-Y, Cheng H-C. A fast and effective system for detection of neonatal jaundice with a dynamic threshold white balance algorithm. Healthcare 2021;9(8):1052. http://dx.doi.org/10.3390/healthcare9081052.

[39] Wu M, Shen X, Lai C, Zheng W, Li Y, Shangguan Z, et al. Detecting neonatal acute bilirubin encephalopathy based on T1-weighted MRI images and learning-based approaches. BMC Med Imaging 2021;21(1):103. http://dx.doi.org/10.1186/s12880-021-00634-z.

[40] Rahayu EP, Widyawati MN, Suryono S. Euclidean distance digital image processing for jaundice detect. IOP Conf Ser: Mater Sci Eng 2021;1108(1):012022. http://dx.doi.org/10.1088/1757-899X/1108/1/012022.

[41] Outlaw F, Nixon M, Odeyemi O, MacDonald LW, Meek J, Leung TS. Smartphone screening for neonatal jaundice via ambient-subtracted sclera chromaticity. PLoS One 2020. http://dx.doi.org/10.1371/journal.pone.0216970.

[42] Daunhawer I. Enhanced early prediction of clinically relevant neonatal hyperbilirubinemia with machine learning. Pediatr Res 2019.

[43] Halder A, Pal SK, Banerjee M, Singh S, Adhikari A, Sarkar PK, et al. A novel whole spectrum-based non-invasive screening device for neonatal hyperbilirubinemia. IEEE J Biomed Health Inform 2019;23(6):2347–53. http://dx.doi.org/10.1109/JBHI.2019.2892946.

[44] Padidar P, Shaker M, Amoozgar H, Khorraminejad-Shirazi M, Hemmati F, Najib KS, et al. Detection of neonatal jaundice by using an android OS-based smartphone application. Iran J Ped 2019;In Press(In Press). http://dx.doi.org/10.5812/ijp.84397.

[45] Kawano S, Zin TT, Kodama Y. A study on non-contact and non-invasive neonatal jaundice detection and bilirubin value prediction. In: 2018 IEEE 7th global conference on consumer electronics. Nara: IEEE; 2018, p. 401–2. http://dx.doi.org/10.1109/GCCE.2018.8574674.

[46] Swarna S, Pasupathy S, Chinnasami B, D. NM, Ramraj B. The smart phone study: Assessing the reliability and accuracy of neonatal jaundice measurement using smart phone application. Int J Contemp Pediatr 2018;5(2):285. http://dx.doi.org/10.18203/2349-3291.ijcp20175928.

[47] Saini N, Kumar A, Khera P. Non-invasive bilirubin detection technique for jaundice prediction using smartphones. IJCSIS 2016;14(8).

[48] Aydın M, Hardalaç F, Ural B, Karap S. Neonatal jaundice detection system. J Med Syst 2016;40(7):166. http://dx.doi.org/10.1007/s10916-016-0523-4.

[49] Bhagat P, Raghuwanshi M, Singh K, Damke S, Quazi S. Development of jaundice detection approaches in neonates. J Univ Shanghai Sci Technol 2020;22(12):918.

[50] WorldMedicalAssociation. World medical association declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013;310(20):2191. http://dx.doi.org/10.1001/jama.2013.281053.

[51] Cover T, Hart P. Nearest neighbor pattern classification. IEEE Trans Inform Theory 1967;13(1):21–7. http://dx.doi.org/10.1109/TIT.1967.1053964.

[52] Stanton JM. Galton, pearson, and the peas: a brief history of linear regression for statistics instructors. J Stat Educ 2001;9(3):3. http://dx.doi.org/10.1080/10691898.2001.11910537.

[53] LeCun Y, Boser BE, Denker JS, Henderson D, Howard RE, Hubbard WE, et al. Handwritten digit recognition with a back-propagation network. In: Proc. Advances in Neural Information Processing Systems. 1990, p. 396–404.

[54] Hua D, Petrina N, Young N, Cho J-G, Poon SK. Understanding the factors influencing acceptability of AI in medical imaging domains among healthcare professionals: A scoping review. Artif Intell Med 2024;147:102698. http://dx.doi.org/10.1016/j.artmed.2023.102698.