Review

Using artificial intelligence to improve body iron quantification: A scoping review

Abdulqadir J. Nashwan [a,b,*], Ibraheem M. Alkhawaldeh [c], Nour Shaheen [d], Ibrahem Albalkhi [e,f], Ibrahim Serag [g], Khalid Sarhan [g], Ahmad A. Abujaber [a], Alaa Abd-Alrazaq [h], Mohamed A. Yassin [i]

[a] Department of Nursing, Hazm Mebaireek General Hospital, Hamad Medical Corporation, Doha, Qatar
[b] Department of Public Health, College of Health Sciences, QU Health, Qatar University, Doha, Qatar
[c] Faculty of Medicine, Mutah University, Al-Karak, Jordan
[d] Faculty of Medicine, Alexandria University, Alexandria, Egypt
[e] College of Medicine, Alfaisal University, Riyadh, Saudi Arabia
[f] Department of Neuroradiology, Great Ormond Street Hospital NHS Foundation Trust, Great Ormond St, London WC1N 3JH, United Kingdom
[g] Faculty of Medicine, Mansoura University, Mansoura, Egypt
[h] AI Center for Precision Health, Weill Cornell Medicine-Qatar, Doha, Qatar
[i] Hematology and Oncology, Hamad General Hospital, Hamad Medical Corporation, Doha, Qatar

ARTICLE INFO

ABSTRACT

This scoping review explores the potential of artificial intelligence (AI) in enhancing the screening, diagnosis, and monitoring of disorders related to body iron levels. A systematic search was performed to identify studies that utilize machine learning in iron-related disorders. The search revealed a wide range of machine learning algorithms used by different studies. Notably, most studies used a single data type. The studies varied in terms of sample sizes, participant ages, and geographical locations. AI's role in quantifying iron concentration is still in its early stages, yet its potential is significant. The question is whether AI-based diagnostic biomarkers can offer innovative approaches for screening, diagnosing, and monitoring of iron overload and anemia.

1. Introduction

1.1. Body Iron metabolism and diseases

Iron is an essential biological element and is strictly regulated at the cellular and systemic levels to avoid both deficiency and excess [1]. Iron homeostasis, often referred to as the iron economy, is predominantly maintained through reabsorption processes [2]. Most of the body's iron—around 20–25 mg/day—is sourced from the recycling of red blood cells in the spleen [2]. In contrast, a mere 1–2 mg/day of new iron is absorbed from the gut to compensate for the iron that is excreted [2]. Disruption in iron hemostasis could lead to various conditions such as iron deficiency which may result from higher iron demand, inadequate external supply, and increased blood loss [3]. Also, iron overload is a serious health issue in both primary (hereditary) and secondary (frequent transfusion and myelodysplasia) hemochromatosis [4]. Iron overload is known to cause damage to multiple organs, such as the liver, heart, and pancreas [5]. Another related disorder is β-thalassemia major, the most prevalent genetic condition globally, which is linked with secondary hemochromatosis [6]. Because thalassemia major is a multiorgan illness, clinical care of afflicted individuals should include iron content testing in the liver, heart, pancreas, spleen, and other organs [7]. The liver is the first and most important site of iron accumulation [8]. Therefore, measuring the liver iron content (LIC) enables us to determine the total body iron content and forecast iron accumulation in other organs. Since various chelator drugs have varied effects on different

---

organs, LIC must be explicitly measured [9].

1.2. Current standard of Care for Quantifying Body Iron

There are many fundamental and technical challenges associated with the current methods of quantifying body iron [10]; such as the low sensitivity and specificity of routine blood tests (e.g., serum iron, serum ferritin, and transferrin saturation … etc.); however, serum iron measurements are frequently employed as an initial screening tool for conditions like anemia or iron overload due to their cost-effectiveness compared to MRI. Modern information technology and computational advancements such as artificial intelligence (AI) present unprecedented opportunities to process vast amounts of diverse data forms and enhance our knowledge about iron homeostasis. This may help clinicians make informed decisions about optimal management strategies for individuals with iron excess or deficiency.

AI could enhance the exploitation of big data and achieve personalized or precision medicine. For example, using image recognition capabilities, machine learning (ML) methods, a subset of AI, can help identify specific histological aspects of chronic conditions, such as chronic liver disease. Although the utilization of AI in medical research and practice is still in the early stages compared to other industries [11], AI-based imaging provides innovative options to predict prognosis and complications, with the eventual goal of precision/personalized medicine [12].

Although the patients with hemoglobinopathies are dispersed across low, middle, and high-income countries, the majority of those residing in low-income countries cannot afford the cost associated with MRI and outsourcing data analysis for iron overload [7]. Over the last few years, medical imaging research has witnessed a transformative shift, largely attributable to the meteoric rise of neural networks. These advanced computational techniques have found profound applications in quantifying body iron, particularly in the domain of T2* Magnetic Resonance Imaging (T2MRI) [36,38]. With their ability to learn complex patterns and process vast amounts of data, neural networks offer a promising alternative [46]. They enhance the accuracy of T2MRI scans in detecting and quantifying iron deposits in body tissues [46]. As a result, the healthcare community is now equipped with more reliable, non-invasive tools to monitor and manage conditions like thalassemia or hemochromatosis, where precise iron quantification is paramount. For instance, Pierre et al. evaluated 1395 MRI datasets submitted for expert manual analysis from 63 scanners on an automated deep-learning-based medical device (DLA R2-MRI) to see its performance in assessing LIC from MRI. Their study revealed that DLA algorithms might offer a solution for inexpensive and dependable LIC patient monitoring worldwide with >90% specificity and sensitivity [13]. However, despite the promising potential of MRI for patient monitoring, its application may prove to be an impractical solution in settings where resources are limited. Moreover, the importance of focusing on low-cost and accessible diagnostics cannot be overstated, as it ensures that vital healthcare services are available to a broader population, including those in resource-limited settings; in this context, the integration of AI could be part of the solution, offering innovative and cost-effective approaches to medical diagnosis and care. Therefore, this scoping review aims to explore how AI can be used to improve the quantification of body iron and its use in the diagnosis and monitoring of related disorders.

2. Methods

The primary objective of this scoping review was to investigate the potential applications of AI in improving the detection of body iron levels. Our methodology rigorously followed the guidelines laid out in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) checklist.

For an article to be considered for inclusion in our review, it needed to satisfy several criteria. Firstly, the article must utilize AI algorithm in

the diagnosis, prognostication, or treatment. Secondly, the study must involve iron quantification or iron-related disorders. Furthermore, it must have been published in English and dated from the year 2010 onward. Finally, the study should not involve any animal research. Notably, there were no limitations regarding the form or nature of the article.

To assemble the widest possible range of relevant literature, we employed a systematic search strategy. This involved using computer-based searches across several key databases, namely MEDLINE, EMBASE, Scopus, IEEE Xplore, and ACM DL. Search terms included those related to target patients, such as iron, hemochromatosis, hemochromatosis, ferrous, ferritin, and ferric, as well as those related to AI, such as "Artificial Intelligence," "Machine Learning," "Deep Learning," "supervised learning," "unsupervised learning," "Ensemble learning," "reinforcement learning," "Decision tree," "K-Nearest Neighbor*," "Support vector machine*," "Recurrent neural network*," "convolutional neural network*," "Artificial neural network*," "Deep Neural Network*," "Naïve Bayes," "Naive Bayes," "Fuzzy Logic," "K-Means," "Random Forest," "Long Short-Term Memory Network*," XGBoost, "Gradient Boost*," adaboost, and "Multilayer Perceptron." The search concentrated on medical and computer science databases (Fig. 1).

Three authors screened the title and abstract of the identified studies independently; in case of conflict, a fourth author was consulted. The articles were chosen for the full-text screening step if they were eligible. To ensure a thorough search, reference lists from included research were manually checked for eligible literature.

Three authors extracted the data shown in (Tables 1, 2) into predefined extraction forms. We extracted the Study ID, Study Design, country, AI model used, study participants, Sample size, Gender (F: M), and Age of participants (mean, SD) for each study. The aim of the study, a summary of methods, and a summary of results were also retrieved. After extracting data from the included research, we used a narrative approach for data synthesis.

We retrieved 2782 studies by searching the identified bibliographic databases. Of those studies, we removed 471 duplicates; we then screened the titles and abstracts of the remaining 2311 studies. The screening process led to the exclusion of 2189 studies. After reading the full texts of the remaining 122 studies, we excluded 102, as they only met some eligibility criteria for the reasons detailed. Thus, we included the remaining 21 studies (Fig. 1).

3. Results

This scoping review identified various studies exploring the use of AI models to detect body iron across different data types. Among the 21 studies, AI techniques based on deep learning models and algorithms were utilized in 6 (28.5%), with convolutional neural networks (CNN) being the most widely used for image analysis, such as MRI scans. Machine learning was used in the other 15 (71.5%) studies for data types like serum measurements, employing methods including Bayesian decision trees, artificial neural networks (ANN) … etc. Fourteen (66.6%) studies used only one source of data to develop their models, while the rest of the studies (7/21, 33.3%) utilized more than one source of data.

The wide range of data types used in the models included demographic data (e.g., patient's age, sex, and ethnicity), clinical data (medical history and family history), laboratory data (patient's blood tests such as red blood cell count, hemoglobin concentration, hematocrit, mean corpuscular volume, ferritin, and transferrin). Techniques like logistic regression, support vector machines (SVM), Extreme Gradient Boosting (XGBoost) … etc. were applied across these various data categories.

The study participants were healthy, suspected, or diagnosed patients with conditions such as iron deficiency anemia (IDA), β-thalassemia, hemochromatosis, and other iron-related diseases. The sample sizes varied across studies, ranging from 31 to 41,764 patients. The gender distribution among participants varied, with some studies
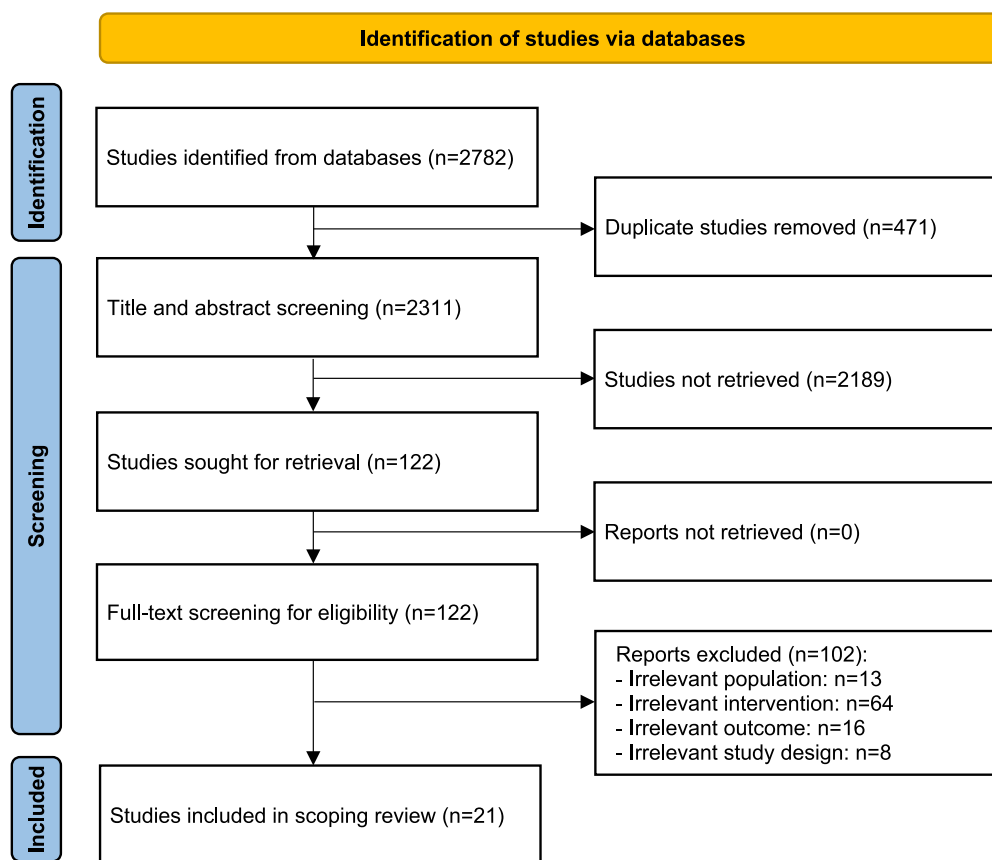
Identification of studies via databases

Identification	
Studies identified from databases (n=2782)	→ Duplicate studies removed (n=471)

Screening	
Title and abstract screening (n=2311)	→ Studies not retrieved (n=2189)
Studies sought for retrieval (n=122)	→ Reports not retrieved (n=0)
Full-text screening for eligibility (n=122)	→ Reports excluded (n=102): - Irrelevant population: n=13 - Irrelevant intervention: n=64 - Irrelevant outcome: n=16 - Irrelevant study design: n=8

Included	
Studies included in scoping review (n=21)	

Fig. 1. PRISMA-ScR flow diagram of the included studies.

reporting male predominance and others reporting female predominance, with a study reporting a female sample. The mean ages of participants ranged from 9.6 to 74 years. The included studies were conducted in 12 countries (Tables 1,2).

Overall, these studies demonstrate the diverse AI approaches employed and the range of patient populations targeted in detecting body iron for the main purposes of disease screening, quantifying iron levels from different organs, detection, and discrimination between iron-related disorders (Table 1). Sorokin et al. (2022) aimed to provide reference ranges for spleen iron concentration and identified genetic associations using genetic analysis methods [14]. Jahangiri et al. (2021) compared different AI models for the differential diagnosis of β-thalassemia trait and iron deficiency anemia, finding that the Bayesian Logit Treed (BLTREED) model outperformed the Classification and Regression Trees (CART) model [15]. Azarkhish et al. (2012) developed an artificial neural network (ANN) model for diagnosing IDA and predicting serum iron levels, with the ANN model achieving high accuracy and sensitivity [16]. Ohara et al. (2021) developed an AI-supported anemia control system (AISACS) that showed high performance in providing dosage directions for erythropoiesis-stimulating agents and iron supplements [17]. Erten et al. (2022) proposed a computer-based model that accurately differentiated IDA and β-thalassemia [18]. Kurstjens et al. (2022) developed machine learning algorithms that accurately predicted low ferritin levels, identifying new iron deficiencies [19]. Terzi et al. (2022) develop machine learning models for IDA, considering various classification issues in IDA, including outliers, class imbalance, the existence of noise, and multicollinearity [34]. Other studies also utilized AI techniques or DLA algorithms, such as the continuous effort for quantification, classification, and automate it of LIC measurements from MRI, that could provide the solution for globally affordable and reliable patient LIC measurements, as proposed by [13,20,22,29,31,33] and ML for discriminating between different iron-related conditions [21,24,26,32].

Finally, Conde et al., 2020 proposed a new risk stratification model for screening hereditary hemochromatosis patients [32]. Overall, the studies highlighted the potential of AI in improving the detection, diagnosis, and management of body iron-related disorders by achieving high accuracy and providing valuable insights into genetic associations and predictive modeling (Table 2).

4. Discussion

In this scoping review, we assessed 21 studies that explored AI's role in body iron quantification across diverse data types. The participants spanned a wide range of conditions and demographics. These investigations underscore AI's potential in iron-related disorder screening, diagnosis, and monitoring, showcasing high accuracy and predictive modeling capabilities. Key studies revealed innovations in differential screening, diagnosis, dosage recommendations, and affordable iron quantification using AI techniques.

4.1. AI and Iron overload

4.1.1. Quantification of Iron overload

Repeated blood transfusions can lead to organ failure and tissue damage due to iron overload [35]. No known mechanism exists for excreting excess iron from the body. Iron accumulation in the body must be carefully watched to guide the response to treatment. Liver iron concentration is a reliable indicator of iron overload and could be assessed noninvasively using MRI [36].

The clinical value of liver biopsy, the conventional method for measuring liver iron content (LIC), is constrained by its invasiveness, expense, and sample variability. Due to the hazards of thrombocytopenia in individuals with iron overload, the biopsy is less preferred and unsuited for repeated treatment monitoring measures. Also, elevated

Table 1

Presents the findings from various studies that employed different AI models for detecting body iron.

Study ID	Study design	Country	AI model used	Participants	Sample size	Gender (F:M)	Age (mean ± SD)
Sorokin et al. 2022 [14]	Cross-Sectional Observational study	UK	Convolutional neural network (CNN) based on the U-Net	Participants from the UK Biobank who have 3D neck-to-knee and the quantitative liver single-slice MRI sequences available	41,764 patients	Female: 51.8% Male: 48.2%	64.2 (7.73)
Jahangiri et al. 2021 [15]	Cross-Sectional Observational study	Iran	Machine learning (Bayesian Decision Tree)	Patients with IDA or βTT	907 patients	Female: 592 (65%) Male: 315 (35%)	25 ± 16.1
Azarkhish et al. 2012 [16]	Cross-Sectional Observational study	Iran	Artificial neural network (ANN) and Adaptive neuro-fuzzy inference system (ANFIS)	Patients suspected of anemia	203 patients	Females: 111 Males: 92	55.8 ± 17.78
Ohara et al. 2021 [17]	Retrospective observational study	Japan	A dense neural network was used for ESAs and a recurrent neural network (RNN) was used for ISs	Hemodialysis patients	NS1 + NS2 + NK1 = 130 + 81 + 16 = 227	Females: 91 Males: 136	Mean age: S1 = 78.0 S2 = 65.6 S3 = 68.3
Erten et al. 2022 [18]	Retrospective cross-sectional analysis	Turkey	Automated ailment classification model consists iterative chi2 (IChi2) feature selection and classification phases using 24 machine learning classifiers	IDA, BTT, and a group of normal individuals	2042 patients	NR	Between 18 and 88 years old
Kurstjens et al., 2022 [19]	Experimental	Netherlands	Random forest classifier model	Anemic patients	• 3797 from the Jeroen Bosch Hospital. • 8021 from Medlon BV. • 191 from St Jansdal Hospital	• Jeroen Bosch Hospital: 48% Male,52% Female • Meldon BV: 49%male, 51%female • St. Jansdal Hospital: 57%male, 43%female	• Mean age= • Jeroen Bosch Hospital: 68 • Meldon BV: 74 • St. Jansdal Hospital: 74
Liu et al. 2020 [20]	Experimental	USA	Modified U-Net style CNN	Patients with hemochromatosis	31 patients	Males: 16 Females: 15	Mean age = 9.6
Ayyıldız et al. 2019 [21]	Experimental - Diagnostic	Turkey	Support Vector Machine (SVM) and K-Nearest Neighbor (KNN)	β-thalassemia and IDA	342 patients	Females: 272 Males: 70	1.5:88
Positano et al. 2023 [22]	Retrospective	Italy	Four deep-learning convolutional neural networks: 1.HippoNet-2D, 2. HippoNet-3D, 3.HippoNet-LSTM, 4.and an ensemble network Hippo Net-Ensemble	Thalassemia major patients	1069 patients	NR	NR
Conde et al. 2020 [23]	Retrospective	Luxemburg (authors), north America (sample)	Machine learning: logistic regression (LR), decision trees (DT), random forests (RF), extreme gradient boosting (XGB), multilayer perceptron (MLP), support vector machine (SVM) and k-nearest neighbors (KNN)	The authors selected 254 cases of hereditary haemochromatosis and 701 controls from The HEIRS cohort, based on their HFE C282Y homozygosity status and iron overload phenotype	955 patients	NR	NR
Yılmaz et al. 2012 [24]	Descriptive	Turkey	Artificial neural network (ANN) models	Women patients	2600 patients	All females (2600)	Over 18
Hennek et al. 2016 [25]	Experimental	USA	Logistic regression after input data from Aqueous multiphase systems (AMPS)	Patients from Boston Children's Hospital	152 patients	Female: 74 Male: 78	Age ≥ 15 yrs. (n = 47) Age ≥ 5 yrs. < 15 yrs. (n = 40)

Table 1 (*continued*)

Study ID	Study design	Country	AI model used	Participants	Sample size	Gender (F:M)	Age (mean ± SD)
							Age < 5 yrs. (n = 65)
Çil et al. 2020 [26]	Comparative	Turkey	Machine learning	Patients diagnosed with β-thalassemia and iron deficiency anemia	342 patients	Females: 272 Males: 70	1–88 years
Martini et al. 2022 [27]	Experimental and retrospective	Italy	Deep convolutional neural network (CNN) with U-Net architecture	Patients with iron overload diseases	210 patients	Females: 106 Males: 104	38.2 ± 12.8 years
Yilmaz et al. 2013 [28]	Descriptive	Turkey	Fuzzy inference system	Patients with IDA	100 patients	NR	NR
Saiviroonporn. et al. 2018 [29]	Retrospective	Italy	Fuzzy c-means clustering (FCM)	Chronically transfused patients with thalassemia major	471 patients	Males: 139 Females: 332	21.7 ± 11.6 years
Pierre et al. 2017 [30]	Experimental	Vietnam	Artificial neural network (ANN)	Thalassemia patients	100 patients	NR	NR
Pierre et al. 2022 -updated 2023 [13,31]	Prospective, validation study	Australia, Turkey, Egypt	Automated deep-learning-based medical device (DLA R2-MRI), that uses convolutional neural network (CNN)	Patients with thalassemia, hereditary hemochromatosis, sickle cell disease, myelodysplastic syndrome (MDS), other, unknown	Thalassemia's (477), hereditary hemochromatosis (168), sickle cell disease (152), MDS (11), other (316), unknown (271) = 1395 patients	NR	NR
Laengsri et al., 2019 [32]	Retrospective	Thailand	Machine Learning	Patients with hypochromic microcytic anemia (HMA) and were diagnosed as reflecting IDA or TT	186 patients	NR	39.15 ± 9.61 years
Wantanajittikul et al. (2021) [33]	retrospective	Thailand	Unsupervised learning algorithm fuzzy c-means clustering combined with anatomical landmark data	Thalassemia major patients	471 patients	Males: 139 Females: 332	21.7 ± 11.6 years
Erol Terzi et al. (2022) [34]	Retrospective	Turkey	Machine learning (XGBOOST)	Cases diagnosed with malaise and fatigue (ICD-10 code: R53). According to laboratory results, they were diagnosed with IDA or were not compatible with IDA	516 patients	NR	46.5 ± 10.3 years (18–89)

ferritin can indicate infection or inflammation, making serologic markers like ferritin and transferrin sensitive but unspecific for detecting iron overload. In addition, Computed tomography (CT) and ultrasound (US) are not useful for measuring liver iron. For the above reasons, MRI is rapidly gaining popularity and acceptance as the most widely used non-invasive diagnostic method for iron overload. It is frequently used to evaluate iron distribution, identify grade, and track the effectiveness of treatments such as chelation therapy [37].

Labranche et al. presented a study highlighting the efficacy of multi-gradient-echo (GRE) acquisitions in effectively quantifying iron, demonstrating particular sensitivity to low levels of iron [37]. Additionally, they noted the advantages of Quantitative Susceptibility Mapping (QSM) reconstructions, which offer higher specificity for iron than R2/R2*. This is due to QSM's ability to separate the diamagnetic component of increased protein content attributed to fibrosis, a condition that can exert a fairly strong R2/R2* effect. However, the authors acknowledged that the benefit of this approach might be minimal for organs without fibrosis, potentially limiting its broader applicability. In addition, using MRI, 102 patients with iron overload and 13 controls were examined with R2 (1/T2) and R2* (1/T2*) techniques by Wood et al. [50]. Both correlated closely with HIC (r2 > or = 0.95), but R2 had a curvilinear relationship. The study found that combined measurements did not improve accuracy; both methods could accurately gauge HIC with appropriate techniques. Moreover, Chavhan et al. revealed that techniques utilizing T2* relaxation can depict hemorrhage, calcification, and iron deposition and calcification in various tissues, with specific settings enhancing these applications [51].

Due to the lack of standardization and different reference levels in MRI-based LIC quantification due to various published techniques, a recent consensus was published [38]. It stated the current evidence that R2- and R2*- based relaxometry approaches and signal intensity ratio (SIR), the three main MRI techniques developed and validated to assess LIC at 1.5 T or 3 T with SIR, do not have regulatory approval [38]. They explained the known mechanisms of these techniques and identified potential confounding variables that may influence the precision and repeatability of MRI results. One such aspect is the coexistence of fat and water in the liver, which might create bias in calculations of the liver iron content using specific MRI techniques. Additionally, fibrosis can make it difficult to estimate iron levels by altering relaxation rates, albeit having a less severe effect than iron excess. Macroscopic magnetic field inhomogeneities close to the liver, especially in particular MRI modalities, might cause signal dephasing and overestimating tissue iron content. The accuracy of R2 relaxometry for iron quantification can be compromised by inhomogeneities in radiofrequency energy deposition (B1 transmission) in the liver, especially at higher field strengths. Finally, noise may cause estimations of liver iron to be skewed, especially at higher iron concentrations [38]. Confounder-corrected R2*-based LIC is the most feasible method with the strongest degree of evidence for accurate and reproducible measurement of LIC, according to the existing literature and the combined knowledge of this consensus panel. When accessible, these techniques are advised as the first-line method for iron quantification. They are commercially available at 1.5 T and 3 T. The use of SIR and R2-based LIC quantification is well supported by moderate to strong levels of evidence [38].

The recent advancements in machine learning were attempted to mostly automate the assessment of LIC using R2 and R2*-based relaxometry. According to the literature, the following stages must be automated before the entire LIC computation from R2* procedure can be

Table 2

Summary the results of various studies related to using artificial intelligence to improve the quantification of body iron.

Study ID	Aim of study	Summary of methods	Summary of results
Sorokin et al. 2022 [14]	1- Provide a reference range for spleen iron concentration. 2- Identify associations between spleen iron levels and genetic.	They performed conditional analysis, fine mapping, colocalization studies, heritability estimates, genetic correlation, exome sequence quality control, rare variant association study, and replication analysis to characterize the genetic signals associated with spleen iron.	1- The average spleen iron was 0.92 ± 0.32 mg/g, significantly lower than liver iron of 1.24 ± 0.29 mg/g, 2- GWAS of spleen iron identified seven loci reaching genome-wide significance, including SLC40A1, MS4A7, SPTA1, and ANK1.
Jahangiri et al. 2021 [15]	1- To compare the Bayesian Logit Treed (BLTREED) model with the Classification and Regression Trees (CART) model for the differential diagnosis of β-thalassemia trait from iron deficiency anemia based on simple laboratory test results. 2- To evaluate the predictive performance and diagnostic accuracy of the BLTREED and CART models 3- To demonstrate the advantages of the BLTREED model over the CART model.	1- Authors measured hematological parameters using a Sysmex KX-21 automated hematology analyzer. 2-They applied (BLTREED) model and (CART) model to differentiate between β-thalassemia trait and iron deficiency anemia based on the hematological parameters.	1- The study found that the BLTREED model had a higher accuracy and Youden's index than the CART model. 2- The study concluded that the BLTREED model is a suitable and helpful method for discriminating between β-thalassemia trait and iron deficiency anemia.
Azarkhish et al. 2012 [16]	1- To develop an artificial neural network (ANN) and an adaptive neuro-fuzzy inference system (ANFIS) to diagnose iron deficiency anemia (IDA) and to predict serum iron level based on four accessible laboratory data (MCV, MCH, MCHC, Hb/RBC). 2- To compare the performance of these models with logistic regression 3- To evaluate the accuracy, sensitivity, specificity, and other measures of these models in	The study developed and trained three models to diagnose iron deficiency anemia (IDA) and predict serum iron level: artificial neural network (ANN), adaptive neuro-fuzzy inference system (ANFIS), and logistic regression.	1- The ANN model achieved the highest accuracy (96.29%), sensitivity (96.8%), specificity (95.6%) in diagnosing IDA compared to ANFIS and logistic regression models. 2- The ANN model predicted the serum iron level with high accuracy and acceptable precision

Table 2 (*continued*)

Study ID	Aim of study	Summary of methods	Summary of results
	diagnosing IDA and predicting serum iron level in patients.		
Ohara et al. 2021 [17]	1- To develop an artificial-intelligence-supported anemia control system (AISACS) that can provide dosage directions for erythropoiesis-stimulating agents (ESAs) and iron supplements (ISs) for hemodialysis patients. 2- To evaluate the performance and validity of AISACS and compare it with the decisions of experienced physicians. 3- To contribute to the improvement of anemia management.	1- The study used a neural network model that learned from the dosage direction data of physicians for ESAs and ISs. 2- The study performed two types of validations: leave one patient out cross-validation (LOPO) and raw data validation (RDV).	1- The study found that AISACS exhibited high performance with correct classification rates of 72%–87% and clinically appropriate classification rates of 92%–98% for ESAs and ISs. 2- The study also found that AISACS sometimes gave better timing than physicians for changing dosage directions.
Erten et al. 2022 [18]	The study aims to develop a machine learning model that can be used to differentiate between iron deficiency anemia (IDA) and beta thalassemia trait (BT).	1. Iterative Chi2 feature selector: This method is used to select the most important features for the classification task. The method starts by selecting the feature with the highest Chi2 value. Then, it iteratively selects the feature with the highest Chi2 value that is not correlated with the previously selected features. This process is repeated until a desired number of features is selected. 2. 24 different classifiers: The paper uses 24 different classifiers to classify the data. The classifiers are a mixture of traditional machine learning algorithms (e.g., decision trees, support vector machines, naive Bayes) and ensemble learning algorithms (e.g., bagging, boosting, random forests). The paper then evaluates the performance of the	The best performing classifiers for IDA and BT were Medium Gaussian Support Vector Machine (MGSVM) and Coarse Tree (CT), respectively. The MGSVM classifier achieved an accuracy of 97.48% on the IDA dataset and 99.73% on the BT dataset. The CT classifier achieved an accuracy of 99.47% on the IDA dataset and 98.57% on the BT dataset. This model may be beneficial for rational laboratory use.

Table 2 (*continued*)

Study ID	Aim of study	Summary of methods	Summary of results
		model on two datasets: a homogenous data set with 159 BT patients and a heterogeneous data set with 1883 IDA and BT patients. They used 10-fold cross-validation and hold-out validation to measure the performance of their model	
Kurstjens et al. 2022 [19]	1- To develop and evaluate a machine learning algorithm that automatically assesses the risk of low body iron storage, reflected by low ferritin plasma levels, in anemic primary care patients. 2- To compare the performance of the algorithm to twelve, who predicted if patients with anemia have low ferritin levels based on laboratory test reports (complete blood count and CRP).	1- The study used laboratory reports from anemic primary care patients. 2- The study developed two machine learning algorithms (one for Siemens and one for Roche analyzers). 3- The study compared the performance of the algorithms. 4- The study implemented the algorithm in the laboratory information system and measured the number of new iron deficiencies identified by the algorithm.	1- The two developed algorithms achieved high accuracy in predicting low ferritin levels 3- The implementation of the algorithm in the laboratory system resulted in one new iron deficiency diagnosis on average per day.
Liu et al. 2020 [20]	The first aim is to explore the use of CNNs to segment livers and exclude vasculature for the purpose of LIC analysis on low resolution pediatric T2* scans. The second aim of this paper is to measure the impact of input channel depth on segmentation accuracy and to see the agreement of LIC calculations with manual methods.	1.They performed Multi echo Gradient Recalled Echo (GRE) MRI sequence for T2* relaxometry was performed for 79 exams on 31 patients with hemochromatosis for iron quantification analysis. 275 axial liver slices were manually segmented as ground truth masks. A batch normalized U-Net with variable input width to incorporate multiple echoes was used for segmentation, using DICE as the accuracy metric. 2. ANOVA was used to evaluate the significance of channel width changes in	ANOVA indicates a significant increase segmentation accuracy over single channel starting at 3 channels. The incorporation of all channels results in an average DICE of 0.86, an average increase of 0.07 over single channel which is a good level of accuracy. The calculated LIC from CNN segmented livers agrees well with manual segmentation. The study's findings suggest that channel width optimization can be used to improve the performance of CNNs for liver and vessel segmentation.

Table 2 (*continued*)

Study ID	Aim of study	Summary of methods	Summary of results
		segmentation accuracy. Linear regression was used to model the relationship of channel width on segmentation accuracy. Liver segmentations were applied to relaxometry data to calculate liver T2* yielding liver iron concentration (LIC) derived from literature-based calibration curves. Manual and CNN based LIC values were compared with Pearson correlation. Bland Altman plots were used to visualize differences between manual and CNN based LIC values.	
Positano et al., 2023 [22]	To develop a deep-learning method for unsupervised classification and staging of LIC from magnitude T2* multiecho MRI.	A dataset of 1069 thalassemia major patients were used. The images were acquired using a 5 T MRI scanner and a T2*-weighted multiecho sequence. The LIC was measured using a standard method. Four deep learning convolutional neural networks (CNNs) were trained on the dataset: HippoNet-2D: A 2D CNN that was trained on magnitude images. HippoNet-3D: A 3D CNN that was trained on the magnitude images. HippoNet-LSTM: A CNN-LSTM hybrid model that was trained on magnitude images. HippoNet-Ensemble: An ensemble model that was trained on the outputs of the HippoNet-2D, HippoNet-3D, and Hippo Net-LSTM models. The CNNs were evaluated using a test set of 210 images. The performance of the models was	The HippoNet-Ensemble model achieved the best performance, with an accuracy of 0.96, a sensitivity of 0.93, and a specificity of 0.97. The other models also performed well, with accuracies ranging from 0.93 to 0.95. The results were reproducible, as the models were able to achieve similar performance on the test set. The results were consistent with previous studies that have shown that deep learning can be used to stage LIC from multiecho MR images.

Table 2 (*continued*)

Study ID	Aim of study	Summary of methods	Summary of results
		assessed by evaluating the accuracy, sensitivity, and specificity of the staging.	
Conde et al. 2020 [23]	The aim of the study was to propose a new disease risk screening for hereditary hemochromatosis (HH). The researchers wanted to develop a more accurate and reliable screening test than the current methods and to identify the most relevant risk factors for HH in the family study of the HEIRS cohort and for this disease.	1. The family and family history datasets from the family study were merged. The data was cleaned and categorical variables with more than two classes were encoded using a one-hot-encoding approach. 2. feature selection based on six different methods (statistical and machine learning-based) was performed and eight different sets of risk factors were manually selected. 3. Last step, each of the selected risk factor sets were evaluated using different machine learning algorithms. First, the data was split into training and testing sets using 10-fold stratified cross-validation (CV). The hyperparameters of each ML algorithm were tuned using Grid Search and tenfold stratified CV and optimized for F1 score. After hyperparameter optimization, the optimal model was trained and evaluated on an unseen test set. This step was repeated 10 times. After final performance estimate, the best model including the best feature set were selected, and hyperparameter optimization was run on the whole dataset using Grid Search and 10-fold stratified CV.	1. The final model was trained on an extreme gradient boosting classifier using the most relevant risk factors: The model was able to achieve an area under the receiver operating characteristic curve (AUC) of 0.94 ± 0.02 which is a good performance. 2. The model was able to outperform the iron overload screening (IRON) tool, which is a commonly used tool for screening individuals at risk of hereditary hemochromatosis.
Ayyıldız et al. 2019 [21]	To perform a differential diagnosis of IDA and β-thalassemia by using more	1- The study used two machine learning algorithms, SVM and KNN, to	1- The study found that the RBC indices showed higher performance for

Table 2 (*continued*)

Study ID	Aim of study	Summary of methods	Summary of results
	comprehensive RBC indices and machine learning techniques including SVM and KNN.	classify the patients into IDA or β-thalassemia groups based on their RBC indices. 2- The study used a feature selection algorithm, NCA, and the features selected by NCA were compared with those used in the literature.	the discrimination of IDA and β-thalassemia. 2- The study found that gender plays a role in the discrimination of IDA and β-thalassemia.
Yılmaz et al. 2012 [24]	To compare the performance and accuracy of different types of artificial neural network models in disease diagnosis.	1- The study used six types of artificial neural network (ANN) models: Feedforward Networks (FFN), Cascade Forward Networks (CFN), Distributed Delay Networks (DDN), Time Delay Networks (TDN), Probabilistic Neural Network (PNN), and Learning Vector Quantization (LVQ) networks. 2- The study compared the performance and accuracy of these models in diagnosing iron deficiency anemia and compared its results with another article that used a similar approach.	1- The study found that Feed Forward Distributed Time Delay network had the highest sensitivity and accuracy in diagnosing iron deficiency anemia in women, with 97.60% sensitivity and 99.16% accuracy. 2- The study also found that Probabilistic Neural Network had the lowest sensitivity and accuracy among the six models
Hennek et al. 2016 [25]	1- To describe a low-cost and rapid method to diagnose IDA using AMPS. 2- To compare its performance with other methods such as hemoglobin concentration and reticulocyte hemoglobin concentration.	1- Training a machine learning algorithm (logistic regression) to discriminate IDA from normal samples and other anemias using the red intensity traces from the scanned images. 2- Evaluating the sensitivity, specificity, and area under the curve (AUC) of the test for different subpopulations and comparing them with other methods.	1- The test can detect IDA by eye with a sensitivity of 84% and a specificity of 78%, 2- The test can slightly improve the diagnosis of IDA using machine learning. 3- The test can also predict hypochromia, micro/hypo anemia, and several red blood cell parameters
Çil et al. 2020 [26]	To compare the performance of different algorithms and indices to distinguish between β-thalassemia and iron deficiency.	1- The study trained and tested five different machine learning algorithms (Logistic Regression, K-Nearest Neighbors, Support Vector	The best accuracy for distinguishing between β-thalassemia and iron deficiency anemia was achieved by Regularized Extreme Learning

Table 2 (*continued*)

Study ID	Aim of study	Summary of methods	Summary of results
		Machine, Extreme Learning Machine and Regularized Extreme Learning Machine). 2- The study evaluated the performance of the classification algorithms using accuracy, sensitivity, specificity, precision, and F1-score metrics.76	Machine with 95.59%.
Martini et al. 2022 [27]	To develop a fully automated method for the regional analysis of myocardial T2* distribution using a deep convolutional neural network (CNN).	1- Generating T2* maps and discarding the last echo time if the fitting error was >5%. 2- Developing a deep convolutional neural network (CNN) with U-Net architecture. 3- Training the CNN model using data augmentation, Adam optimizer, and weighted cross-entropy loss function for 2000 epochs. 4- Evaluating the segmentation accuracy and T2* values obtained by the CNN model.	The CNN model showed high sensitivity (97.6%), specificity (100%), accuracy (98.8%), and intraclass correlation coefficient (0.996) for detecting myocardial iron overload using a threshold of 20 ms for global T2 values.
Yilmaz et al. 2013 [28]	To introduce an application of fuzzy expert system for the diagnosis of iron deficiency anemia.	1- The authors designed a fuzzy expert system for the diagnosis of iron deficiency anemia using five input parameters: Hb amount, MCV, SI, TIBC and Ferritin amount. 2- The authors used Delphi 2010 as a visual programming language to design and run the system.	1- The authors detected the success level of their system between 90% and 95.8%. 2- The authors concluded that their system is convenient and accurate for the diagnosis of iron deficiency anemia.
Saiviroonporn et al. 2018 [29]	To develop an automated segmentation method for R2* iron-overloaded liver images using a fuzzy C-mean clustering scheme.	1.The study utilized R2* MRI images of iron-overloaded livers. 2.A fuzzy C-mean clustering algorithm was employed for image segmentation. 3.The algorithm used intensity values and spatial information to classify image pixels into different clusters.	1.The proposed segmentation method achieved accurate and reliable segmentation of R2* iron-overloaded liver images. 2.The fuzzy C-mean clustering scheme effectively separated liver regions from background and other tissues. 3.
		4.The algorithm was optimized by adjusting parameters, including the number of clusters and fuzziness index. 5.The segmentation method was evaluated using a quantitative assessment, comparing the results with manual segmentations performed by experts.	The method demonstrated robustness and consistency across different datasets. 4.Quantitative evaluations showed high similarity and agreement between the automated segmentation and manual segmentations performed by experts. 5.The automated segmentation approach provided a time-efficient alternative to manual segmentation, enabling efficient analysis of large datasets.
Pierre et al. 2017 [30]	1- To determine the limits of agreement between measurements of liver iron concentration by a reference standard spin-density-projection-assisted (SDPA) R2-MRI method (FerriScan ®) and a method using trained artificial neural networks (ANNs) to analyze image data. 2- To test the potential of the ANN method as a fast and robust method of analyzing MR image data to generate liver iron concentration values.	1- Image data were acquired using the FerriScan ® protocol. 2- Images were processed by two methods: (1) using the SDPA R2-MRI method and (2) by using a trained artificial neural network (ANN) to generate LIC values. 3- The technician was blinded from the SDPA R2-MRI results and vice versa. 4- The data were analyzed using the statistical methods of Bland and Altman.	1- The ANN method showed a small but statistically significant systematic underestimation of liver iron concentration compared to the SDPA R2-MRI method. 2- The ANN method had high positive predictive values and negative predictive values for predicting liver iron concentration above the clinically relevant thresholds of 7 and 15 mg Fe/g dw.
Pierre et al. 2022-updated 2023 [13,31]	This study aimed to assess the diagnostic performance and repeatability of a deep-learning-based medical device (DLA R2-MRI) for quantifying LIC from MRI in comparison with the reference standard of the expert manual analysis using spin-density projection assisted (SDPA) R2-MRI	Data collection: The MRI datasets were collected from 63 different scanners between August 2017 and July 2020. The datasets were submitted for expert manual analysis using spin-density projection assisted (SDPA) R2-MRI, which is the reference standard for measuring liver iron concentration (LIC). Data analysis: The	1. Sensitivities and specificities of the automated DLA R2-MRI system for predicting LIC values using the SDPA R2-MRI method above clinically relevant thresholds (3.0, 5.0, 7.0, and 15.0 mg Fe/g dry tissue). Although the bias between DLA R2-MRI and SDPA R2-MRI, the DLA R2-MRI method has acceptable sensitivities and

(*continued on next page*)

Table 2 (*continued*)

Study ID	Aim of study	Summary of methods	Summary of results
		automated measurements of LIC from the DLA R2-MRI device were compared to the manual measurements from the SDPA R2-MRI. The bias and limits of agreement between the two methods were assessed. In addition, the diagnostic performance of the DLA R2-MRI device was assessed using sensitivity and specificity analysis. Repeatability: The repeatability of the DLA R2-MRI device was assessed by recruiting 60 participants with informed consent (50 patients and 10 healthy controls). Each participant was measured twice with the DLA R2-MRI device. The limits of agreement, bias, and repeatability coefficients were determined using Bland Altman statistics.	specificities for predicting LIC results above clinically significant thresholds. However, due to the identified bias, the automated and manual procedures should not be employed interchangeably. 2. 95% of the repeat measures of LIC by DLA R2-MRI fell within specific ratio ranges above and below 3 mg Fe/g dry tissue. The repeatability of the DLA R2-MRI method for LIC measurement is significantly better than liver biopsy methods.
Laengsri et al. 2019 [32]	To develop a computational model for discriminating between iron deficiency anemia and thalassemia trait.	1- The study applied five machine learning techniques (k-NN, DT, RF, ANN and SVM) to construct discriminant models based on seven red blood cell parameters. The models were evaluated using 5-fold cross-validation. The performance of the models was compared with existing discriminant formulas and indices. 2- The study implemented a web-based tool called ThalPred using the SVM model.	The study showed that the SVM model outperformed the other machine learning techniques and the existing discriminant formulas and indices in terms of accuracy, sensitivity, specificity, Matthew's correlation coefficient, Youden's index. The SVM model achieved an external accuracy of 95.59% and an independent accuracy of 96.08%.
Wantanajittikul et al. 2021 [33]	Previous works have proposed automated	1. The authors used a dataset of 553 MR	1.The proposed method was able to produce good

Table 2 (*continued*)

Study ID	Aim of study	Summary of methods	Summary of results
		methods for excluding vessels from the liver region. However, these methods still require the user to manually define the liver region. The aim of this study was to develop an automated liver region segmentation technique to automate the whole process of median LIC calculation.	examinations from 471 thalassemia major patients. 2. The LIC maps were calculated from the MR images and used as the input for the segmentation procedures. 3. The anatomical landmark data were used to restrict the region of interest (ROI) for the segmentation. 4. The liver region was segmented using fuzzy c-means clustering and morphological processes were applied to reduce segmentation errors. 5. The segmentation results were evaluated by comparing them with manual segmentation performed by a board-certified radiologist.	grade output in approximately 81% of all data. 2.Approximately 11% of all data required an easy modification step. 3.The rest of the output, approximately 8%, was an unsuccessful grade and required manual intervention by a user. 4.The proposed method was more accurate and reliable than manual segmentation. 5.The proposed method could be used to automate the liver segmentation process in LIC maps.
Terzi et al. 2022 [34]	So far, studies have investigated a variety of classification issues in IDA, including outliers, class imbalance, the existence of noise, and multicollinearity. However, most datasets are plagued by more than one of these issues. This work aimed to develop several systems that can distinguish between diseased and healthy persons and find the variables that have a substantial effect on these diseases while considering these influential classification challenges.	Data preprocessing: Outliers were detected and removed using Z-score, Relative density-based outlier factor (RDOS), and natural outlier factor (NOF). Feature selection was performed in two stages: Boruta was used to identify significant variables before cross-validation. The training set was transformed to have zero mean and one variance. Class imbalance and noise problems were solved using oversampling (SMOTE) and noise detection-undersampling (EF). Model development: XGBoost was used to develop classification models. Hyperparameters for XGBoost were	When all reported concerns were taken into consideration, the models we obtained were successful in more measures. In order of importance, the XGBoost approach discovered Fe, Ferritin, and Hb features to be essential for all cases, with ferritin being the most important. Feature importance did not differ by the method. The performance findings varied among datasets, with SMOTE and EF + SMOTE outperforming all others in most parameters.

Table 2 (*continued*)

Study ID	Aim of study	Summary of methods	Summary of results
		determined using grid search to maximize Matthew's correlation coefficient performance, taking class imbalance into account. Model evaluation: The models were evaluated using 10-fold cross-validation. They created and tested four different data sets according to classification issues. Performance Metrics. confusion matrix and ROC and related metrics were used.	

carried out automatically. They are listed as follows: There are six steps in the process: 1. manually define an ROI to exclude non-body part; 2. calculate R2* by pixel-wise method with offset model and conversion to LIC map; 3. construct an ROI for the entire liver; 4. manually exclude vessels from the liver parenchyma; and 5. generate a LIC report. 6. The classification and staging of liver iron excess overload [22,33] (Fig. 2). As many centers have limited imaging data, quality control of T2* image capture and image analysis methodologies is challenging. As a result, an automated method can be helpful as a "second reader opinion" or as a

replacement for an external service's diagnosis. Also, it can reduce the average processing time and the workload [22].

Previous work, such as Saiviroonporn et al. [29], proposed methods for automatically excluding main vessels in the user-defined liver regions or automated on step number 4. Also, Liu et al. [20] applied a batch-normalized U-Net with variable input width to incorporate multiple echoes for liver and vessel segmentation in liver iron quantification. For steps 1 and 3, Wantanajittiku et al. [33] proposed a method of fuzzy c-means (FCM) clustering algorithm and anatomical landmark data. In post-processing, morphological techniques were used to reduce segmentation mistakes and modify the liver area for clinical use. The experimental findings indicated that the suggested strategy could boost the effectiveness of traditional FCM clustering. It offered good grade outputs with good assessment matrices for about 81% of all the data. A simple modification step was needed for 11% of the data to fix the segmentation findings. The remaining 8% required manual segmentation. Their trials revealed a strong correlation in the median LIC between our suggested method and the existing method. Intending to take a further step, Positano et al. [22] used ensemble deep-learning CNN networks to automate the staging of LIC in thalassemia major patients, showing a good diagnostic value.

When it comes to R2 relaxometry, Pierre et al. [30] conducted a study to determine the limits of agreement between measurements of LIC by a reference standard spin-density-projection-assisted (SDPA) R2-MRI method (Ferri-Scan) and a method using trained ANNs to analyze image data in thalassemia patients. The data indicate that the ANN measurement method of LIC can be used as a fast and robust method of analyzing MR image data to generate LIC values. As a continuation of their work, they conducted a study to assess the diagnostic performance and repeatability of a DLA R2-MRI for quantifying LIC from MRI in comparison with the reference standard of the expert manual analysis using SDPA R2-MRI. This study aimed to assess the diagnostic performance and repeatability of a DLA R2-MRI for quantifying LIC from MRI in comparison with the reference standard of the expert manual analysis using SDPA R2-MRI; they did it prospectively on 1395 eligible

The interplay between MRI-based liver iron concentration quantification and machine learning.
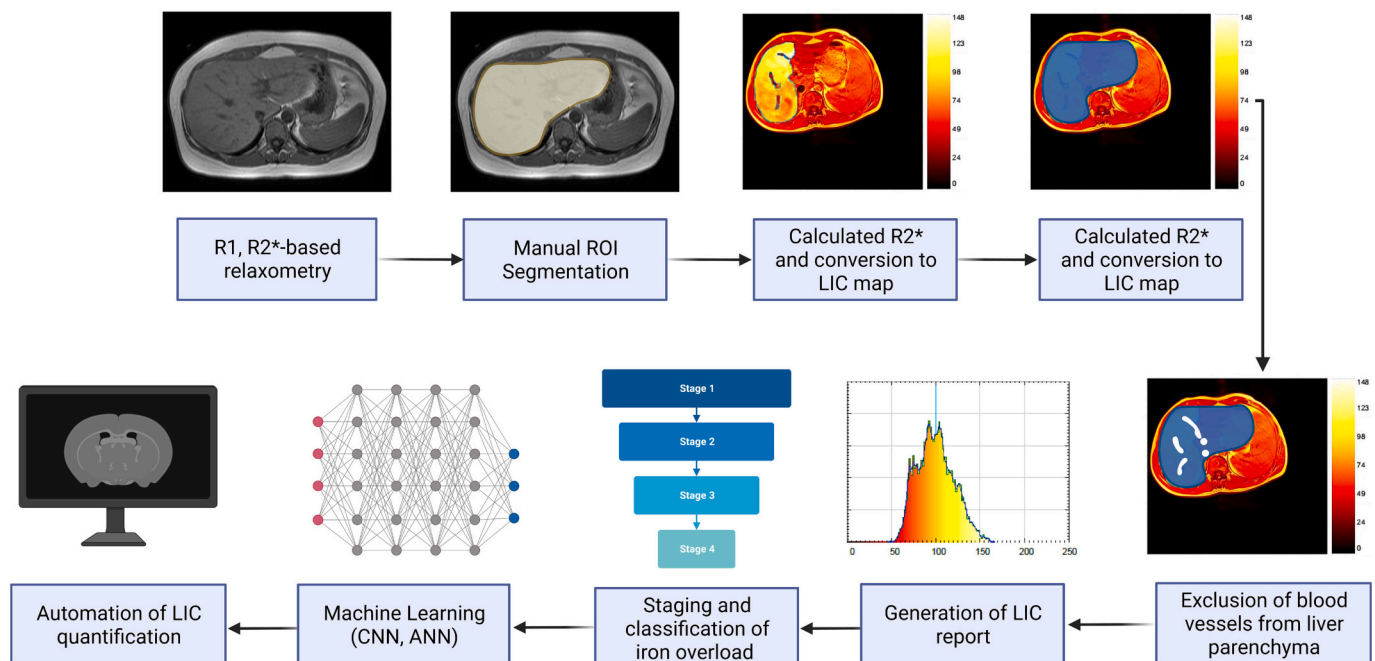

Fig. 2. The interplay between MRI-based LIC concentration quantification and machine learning.

consecutive MRI datasets from 63 different scanners Their study revealed that DLA algorithms might offer a solution for inexpensive and dependable LIC patient monitoring worldwide with >90% specificity and sensitivity. Also, it showed significantly better repeatability compared with liver biopsy methods [13,31].

Although recent attempts to incorporate ML models into automating the process show promise, the implementation in clinical practice is still in its infancy. This is due to a range of limitations, including human-related factors such as the necessity for extensive training and the 'black box' phenomenon of ML. There is also a pressing need for multicentered studies and external validation across diverse clinical settings, as well as a call for globally representative populations. Many current studies suffer from a lack of representative populations, as they often involve only single disease populations, leading to heterogeneous data. Methodology-related model limitations and previously mentioned confounders, such as potential variations between MRI scanners and laboratory reference values, further complicate matters. In light of these challenges, it is particularly crucial to discuss the lack of large training datasets for quantifying organ iron from MRI, especially for populations that are globally representative. For instance, the UK Biobank offers a vast cohort with multiple MRI acquisitions [52], but it is limited mainly to participants of European ancestry. Addressing these issues may pave the way for more universally applicable models.

4.1.2. Screening of hereditary hemochromatosis

Using information from the HEIRS cohort, Conde et al. [23] conducted a study to create a screening test that is better and more trustworthy than the ones now used and to determine the most important risk factors for hereditary hemochromatosis (HH). Based on their HFE C282Y homozygosity status and iron overload phenotype, they chose 254 hereditary hemochromatosis (HH) cases and 701 controls. The most important risk factors, such as age, HFE C282Y homozygosity, iron level, transferrin saturation, serum ferritin level, unsaturated iron-binding capacity, and other variables, were used to train the final model on an extreme gradient boosting classifier. The area under the receiver operating characteristic curve (AUC) for the model was 0.94 0.02, which is regarded to be a good performance. Their risk stratification model outperformed the iron overload screening (IRON) tool, a commonly used tool for screening individuals at risk of hereditary hemochromatosis. However, the above model needs to be validated by clinical experts before transforming it into an easy-to-use application.

4.1.3. Iron overload cardiomyopathy: Diagnosis and prognosis

AI-derived feature tracking analysis in cardiology has maximized the speed and the amount of information harvesting [39]. Echocardiography detects early myocardial dysfunction associated with iron overload, which is usually measured by T2* in transfusion-dependent several types of anemias. For instance, Alonso-Fernandez-Gatta et al. used an AI-derived predictive model to investigate the use of myocardial magnetic resonance feature tracking (MR-FT) to diagnose and predict the prognosis of cardiovascular events in patients with myelodysplastic syndromes [40].

Choi stated that the "cherry on top" of cardiac magnetic resonance as a one-stop modality seems to be feature tracking analysis. However, the limitations associated with this approach may be solved by sharing analysis methods and applying innovative ways to reduce picture collection time [41].

In addition, Martini et al. [27] developed a fully automated method for the regional analysis of myocardial T2* distribution using a deep convolutional neural network (CNN). The CNN model showed high sensitivity (97.6%), specificity (100%), accuracy (98.8%), and intraclass correlation coefficient (0.996) for detecting myocardial iron overload using a threshold of 20 ms for global T2 values. This technique could be successfully applied in the clinical setting for a quick introduction of new T2* CMR sites, particularly in areas of the world with financial limits, and for faster, more accurate, and quality-controlled analysis in

trained centers.

4.2. AI and Iron-deficiency Anemia

4.2.1. An IA approach to Anemia prevalence analysis

According to the World Health Organization (WHO), both developed and developing countries suffer from anemia. The most vulnerable groups are children aged under five years, adolescents, and pregnant women [42]. Anemia affects 42% of children aged five years or less and 40% of pregnant women worldwide [43]. As teens grow up, they are at higher risk of developing health problems that last into adulthood, negatively impacting their quality of life [44]. Anemia occurs when the blood contains insufficient red blood cells (RBCs) or hemoglobin. A person with anemia often experiences dizziness, weakness, fatigue, and shortness of breath. The severity of signs and symptoms depends on factors such as age, gender, pregnancy status, smoking habits, location, and optimal hemoglobin concentration. There are many causes of anemia, including nutritional deficiencies (e.g., iron, vitamin B12, folate, and vitamin A), infectious diseases (e.g., tuberculosis, malaria, human immunodeficiency virus (HIV), parasitic infections), and other disorders (e.g., hemoglobinopathies). Globally, the most prevalent type of anemia is IDA, especially in Asia and sub-Saharan Africa [43,44]. Iron is vital for sustaining cell function and structure in the body, as well as assisting oxygen transport, in addition to being an essential micronutrient [3].

Large datasets can be analyzed, and ML algorithms can perform predictive analytics faster. To handle classification problems, medical statistics traditionally rely on logistic regression (LR) [46]. Public health research uses ML algorithms to solve classification problems using model-free data analytics [47]. Most anemia risk analysis techniques are based on accurate disease prediction. Furthermore, pattern recognition and predictive analytics are becoming increasingly popular in health and medical research [48].

4.2.2. Anemia classification using ML

In some cases, it is critical to appropriately define anemia due to potential implications, such as distinguishing between anemia and β-thalassemia. In the event of a misdiagnosis, patients with β-thalassemia may receive unnecessary iron supplementation. Furthermore, when a β-thalassemia patient is misdiagnosed with IDA, he or she may have children with β-thalassemia or carry it in marriages [26].

Zhao et al. [49] proposed a method for identifying anemia in a general clinical setting. Their model was constructed using four machine learning algorithms: artificial neural networks, support vector machines, naive Bayes, and ensemble decision trees. The models were developed based on 1663 samples and 25 attributes, including hemogram data, age, gender, chronic diseases, and symptoms gathered from the medical records of a Turkish teaching hospital. Its purpose was to provide decision support to medical students and assist medical consultants in making informed decisions. Classifying the data using the four different algorithms achieved a satisfactory success ratio for each algorithm. Bagged Decision Trees (aggregating the output from all the trees and forecasting the outcome) had the highest accuracy (85.6%), followed by boosted Trees (83%) and neural networks (79.6%) [49].

Azarkhish et al. [16] developed an ML model based on 4 common blood data to diagnose IDA(15). The results showed that neural networks (ANN) outperform adaptive neuro-fuzzy inference systems (ANFIS) and logistic regressions in diagnosing IDA. Additionally, the results demonstrated that ANNs had higher accuracy and acceptable precision for predicting serum iron levels in patients with IDA (up to 97% vs. 93% for the logistic regression model). The study developed ANN, ANFIS, and logistic regression models to diagnose IDA based entirely on routine and inexpensive laboratory data. Patients' serum iron levels were also predicted using ANNs and linear regression. Several accessible variables were used to construct the ANN but were not directly diagnostic of IDA. The designed model successfully diagnosed IDA in patients based on routine laboratory data. Using this model, we

can obtain excellent accuracy, non-invasive, affordable, and quick results, which can be applied in clinical practice to guide decision-making [16].The utilization of AI has led to an increased interpretation potential from simple lab tests. This offers significant advantages to developing countries with limited capabilities. Jahangiri et al. [15] showed how a complete blood count (CBC) could differentiate iron deficiency anemia from β-Thalassemia. The authors also visualized the splitting rules of the tree-based machine learning models to further understand the decision pathway. Both models in the study used hemoglobin (Hb), mean corpuscular volume (MCV), and mean corpuscular hemoglobin (MCH) to finally classify the anemia. The AUC of the trained models ranged from 0.94 to 0.98.

Moreover, Laengsri et al. [32] developed a web-based prediction tool based on SVM (AUC = 96.0) for discriminating thalassemia trait and iron deficiency anemia, as well as demonstrated how ML could provide interpretable rules interpretable rules extracted from RF (AUC = 94.6) were provided to represent criteria of each CBC parameter for discrimination. They also mentioned that the two differential diagnoses are close and can coexist. Previous research has shown that IDA occurs frequently in TT. Interestingly, this discovery suggests future research to develop a new efficient model for distinguishing three groups, including TT, IDA, and the coexistence of these two conditions.

When looking at classification issues in IDA, Terzi et al. [34] investigated these challenges and proposed a comprehensive approach to address them. They used a variety of methods, including outlier detection, feature selection, multicollinearity consideration, class imbalance handling, and noise removal, to develop accurate and reliable models for the early diagnosis of IDA, since most datasets are plagued by more than one of these issues. The results showed that the proposed approach was effective in most cases, with the most important features being Fe, Ferritin, and Hb using XGBoost approach with ferritin being the most important. The performance of the models varied depending on the dataset, but overall, they were able to achieve good performance. Kurstjens et al. [19] highlighted the benefits of AI application in clinical practice, demonstrating that the integration of a machine learning algorithm to predict low ferritin levels in primary care patients is a valuable diagnostic tool that can support physicians and specialists in laboratory medicine, and can automatically identify unrecognized iron deficiencies. In their study, laboratory medicine specialists were less accurate in predicting low ferritin concentrations than algorithms, even when they knew the output of the algorithms as a support tool. also, it took less time for each patient. The algorithm's use in the laboratory system resulted in one new iron deficiency diagnosis per day on average.

The current landscape of iron-related disorder screening, diagnosis, and monitoring faces several challenges stemming from the limitations of conventional methods. Routine blood exhibit low sensitivity and specificity, hindering accurate disease detection [53]. Invasive techniques like liver biopsy, while informative, pose substantial risks and discomfort to patients. Furthermore, the accessibility and affordability of advanced devices remain restricted. The applications outlined in this review highlight the potential role of AI in addressing those limitations.

The heterogeneity in the results obtained can be attributed to several reasons. The datasets' diversity, including sources and features, impacts model performance. Furthermore, feature selection, preprocessing, and hyperparameter tuning also shape outcomes. Studies that involve multi-center collaboration often led to the development of stable and generalizable models, in contrast to single-center studies that frequently suffer from limited sample sizes and poorer performance during external and temporal validation. Finally, a collaboration between data scientists and clinicians is required for healthcare ML initiatives to succeed. Understanding how these roles interact and where conflict might occur because of communication challenges, incentives, and conflicting viewpoints is crucial, this would support the future acceptance of current AI applications among physicians.

5. Conclusion

Several imaging and laboratory techniques have been developed to assess iron distribution and severity and monitor treatment response. However, despite and the current limitations and the fact that AI's adoption in assessing iron concentration is still in the early phases, it has great potential to revolutionize the current practice.

Future considerations

To move this field forward, progress in AI for managing body iron levels requires larger, more diverse studies and the integration of various data types with more features, including genetic, clinical, and lifestyle factors. External validation in various settings is a cornerstone for some promising models before being trustworthy for application in clinical practice. This will improve the AI model's accuracy and generalizability. Developing standardized AI model protocols through international collaborations is essential, as is addressing ethical considerations like data privacy and AI decision-making consequences. These advancements will leverage AI's potential in managing body iron-related disorders and propel precision medicine forward.

Practice points

- Artificial intelligence, especially deep learning models and algorithms like convolutional neural networks, shows promising potential in diagnosing and managing iron-related disorders, from iron deficiency to hemochromatosis.
- Current studies in the iron-related disorder domain focus on radiological images. Expanding AI models to integrate various data types—such as genetic markers associated with thalassemia, clinical symptoms of iron overload, or lifestyle factors influencing iron deficiency—can enhance diagnostic precision.
- AI models in this field must be trained and validated on large, diverse patient populations, reflecting the wide spectrum of iron-related disorders. Including diverse age groups, ethnic backgrounds, and geographic locations can lead to more robust models.
- Implementing standardized protocols specific to iron-related disorders for AI model development, training, and validation can foster reliable and generalizable tools. Creating benchmarks for different iron disorders can facilitate consistent evaluations and comparisons.
- Data privacy in the context of genetic and personal health information is paramount. Ensuring ethical deployment and transparent AI decision-making is essential for patient trust in tools managing sensitive iron-related health data.
- AI applications in iron disorder management could act as a complementary "second reader opinion" or even replace some diagnostic services. Tailoring these applications to specific iron-related conditions can reduce processing time and workload, benefiting both clinicians and patients in this specific area.

Research agenda

- Development of AI models for predicting the recurrence of iron-related disorders.
- Evaluation of the clinical value of AI in quantifying body iron levels.
- Investigation of the relationship between different physiological factors and iron-related disorders using AI.
- Creation of a comprehensive AI-driven diagnostic tool for assessing overall body iron status.
- External validation is a cornerstone of the current models and for future proposed studies to have good repeatability and generalizability.

Funding sources

CRediT authorship contribution statement

Abdulqadir J. Nashwan: Conceptualization, Writing – original draft, Writing – review & editing. **Ibrahim M. Alkhawaldeh:** Writing – original draft, Writing – review & editing. **Nour Shaheen:** Writing – original draft, Writing – review & editing. **Ibrahem Albalkhi:** Writing – original draft, Writing – review & editing. **Ibraheem Serag:** Writing – original draft, Writing – review & editing. **Khalid Sarhan:** Writing – original draft, Writing – review & editing. **Ahmad A. Abujaber:** Writing – original draft, Writing – review & editing. **Alaa Abd-Alrazaq:** Writing – original draft, Writing – review & editing. **Mohamed A. Yassin:** Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors have no conflicts of interest to declare.

Acknowledgment

References

[1] Camaschella C, Nai A, Silvestri L. Iron metabolism and iron disorders revisited in the hepcidin era. Haematologica. 2020;105(2):260–72.

[2] Ganz T. Systemic iron homeostasis. Physiol Rev 2013;93(4):1721–41.

[3] Muñoz M, García-Erce JA, Remacha ÁF. Disorders of iron metabolism. Part II: iron deficiency and iron overload. J Clin Pathol 2011;64(4):287–96.

[4] Nashwan AJ, Yassin MA, Mohamed Ibrahim MI, Abdul Rahim HF, Shraim M. Iron overload in chronic kidney disease: less ferritin, more T2*MRI. Front Med (Lausanne) 2022;9:865669.

[5] Ramm GA, Ruddell RG. Hepatotoxicity of iron overload: mechanisms of iron-induced hepatic fibrogenesis. Semin Liver Dis 2005;25(4):433–49.

[6] Galanello R, Origa R. Beta-thalassemia. Orphanet J Rare Dis 2010 May 21;5:11.

[7] De Sanctis V, Kattamis C, Canatan D, Soliman AT, Elsedfy H, Karimi M, et al. β-Thalassemia distribution in the Old World: an ancient disease seen from a historical standpoint. Mediterr J Hematol Infect Dis 2017;9(1):e2017018.

[8] Nashwan AJ, Yassin MA, Abd-Alrazaq A, Shuweihdi F, Abdul Rahim HF, Shraim M. The prevalence of cardiac and hepatic iron overload in patients with kidney failure: a protocol for systematic review and meta-analysis. Health Sci Rep 2022;5(4):e692.

[9] Angelucci E, Brittenham GM, McLaren CE, Ripalti M, Baronciani D, Giardini C, et al. Hepatic iron concentration and total body iron stores in thalassemia major. N Engl J Med 2000;343(5):327–31.

[10] Hernando D, Levin YS, Sirlin CB, Reeder SB. Quantification of liver iron with MRI: state of the art and remaining challenges. J Magn Reson Imaging 2014;40(5): 1003–21.

[11] Abujaber AA, Nashwan AJ, Fadlalla A. Enabling the adoption of machine learning in clinical decision support: a Total interpretive structural modeling approach. Informatics in Medicine Unlocked 2022;33:101090.

[12] Dana J, Venkatasamy A, Saviano A, Lupberger J, Hoshida Y, Vilgrain V, et al. Conventional and artificial intelligence-based imaging for biomarker discovery in chronic liver disease. Hepatol Int 2022;16(3):509–22.

[13] St Pierre TG, Aydinok Y, El-Beshlawy A, Bayraktaroglu S, Ibrahim A, Hamdy M, et al. The diagnostic accuracy and repeatability of an artificial intelligence-based system to obtain automated liver iron concentration measurements using magnetic resonance imaging. HemaSphere. 2023 Apr;7(S1):3–4.

[14] Sorokin EP, Basty N, Whitcher B, Liu Y, Bell JD, Cohen RL, et al. Analysis of MRI-derived spleen iron in the UK biobank identifies genetic variation linked to iron homeostasis and hemolysis. Am J Hum Genet 2022;109(6):1092–104.

[15] Jahangiri M, Rahim F, Saki N, Saki Malehi A. Application of Bayesian decision tree in hematology research: differential diagnosis of β-thalassemia trait from Iron deficiency Anemia. Comput Math Methods Med 2021;2021:6401105.

[16] Azarkhish I, Raoufy MR, Gharibzadeh S. Artificial intelligence models for predicting iron deficiency anemia and iron serum level based on accessible laboratory data. J Med Syst 2012;36(3):2057–61.

[17] Ohara T, Ikeda H, Sugitani Y, Suito H, Huynh VQH, Kinomura M, et al. Artificial intelligence supported anemia control system (AISACS) to prevent anemia in maintenance hemodialysis patients. Int J Med Sci 2021;18(8):1831–9.

[18] Erten M, Tuncer T. Automated differential diagnosis method for iron deficiency anemia and beta thalassemia trait based on iterative Chi2 feature selector. Int J Lab Hematol 2022 Apr;44(2):430–6.

[19] Kurstjens S, de Bel T, van der Horst A, Kusters R, Krabbe J, van Balveren J. Automated prediction of low ferritin concentrations using a machine learning algorithm. Clin Chem Lab Med 2022;60(12):1921–8.

[20] Liu M, Vanguri R, Mutasa S, Ha R, Liu Y-C, Button T, et al. Channel width optimized neural networks for liver and vessel segmentation in liver iron quantification. Comput Biol Med 2020;122:103798.

[21] Ayyıldız H, Arslan Tuncer S. Determination of the effect of red blood cell parameters in the discrimination of iron deficiency anemia and beta-thalassemia via neighborhood component analysis feature selection-based machine learning. Chemom Intel Lab Syst 2019;103886.

[22] Positano V, Meloni A, Santarelli MF, Pistoia L, Spasiano A, Cuccia L, et al. Deep learning staging of liver iron content from multi-echo MR images. J Magn Reson Imaging 2023;57(2):472–84.

[23] Martins Conde P, Sauter T, Nguyen T-P. An efficient machine learning-based approach for screening individuals at risk of hereditary hemochromatosis. Sci Rep 2020;10(1):20613.

[24] Yılmaz Z, Bozkurt MR. Determination of women's iron deficiency anemia using neural networks. J Med Syst 2012;36(5):2941–5.

[25] Hennek JW, Kumar AA, Wiltschko AB, Patton MR, Lee SYR, Brugnara C, et al. Diagnosis of iron deficiency anemia using density-based fractionation of red blood cells. Lab Chip 2016;16(20):3929–39.

[26] Çil B, Ayyıldız H, Tuncer T. Discrimination of β-thalassemia and iron deficiency anemia through extreme learning machine and regularized extreme learning machine based decision support system. Med Hypotheses 2020 May;138:109611.

[27] Martini N, Meloni A, Positano V, Latta DD, Keilberg P, Pistoia L, et al. Fully automated regional analysis of myocardial T2* values for iron quantification using deep learning. Electronics. 2022;11(17):2749.

[28] Yilmaz A, Dagli M, Allahverdi N. A fuzzy expert system design for iron deficiency anemia. In: 2013 7th International Conference on Application of Information and Communication Technologies. IEEE; 2013. p. 1–4.

[29] Saiviroonporn P, Korpraphong P, Viprakasit V, Krittayaphong R. An automated segmentation of R2* Iron-overloaded liver images using a fuzzy C-mean clustering scheme. J Comput Assist Tomogr 2018;42(3):387–98.

[30] St Pierre T, Cooper T, Trang NN, Ha NT, Ton DT, Viet NK, et al. Measurement of liver Iron concentration in a population of non-transfusion dependent thalassemia patients using a trained artificial neural network to analyse magnetic resonance images. Blood. 2017 Dec 8;130:2212.

[31] St Pierre T, Aydinok Y, El-Beshlawy A, Bayraktaroglu S, Ibrahim A, Hamdy M, et al. P1505: using artificial intelligence neural networks to obtain automated liver iron concentration measurements using magnetic resonance imaging – a multi-scanner validation study. HemaSphere. 2022;6:1386–7.

[32] Laengsri V, Shoombuatong W, Adirojananan W, Nantasenamat C, Prachayasittikul V, Nuchnoi P. ThalPred: a web-based prediction tool for discriminating thalassemia trait and iron deficiency anemia. BMC Med Inform Decis Mak 2019;19(1):212.

[33] Wantanajittikul K, Saiviroonporn P, Saekho S, Krittayaphong R, Viprakasit V. An automated liver segmentation in liver iron concentration map using fuzzy c-means clustering combined with anatomical landmark data. BMC Med Imaging 2021;21 (1):138.

[34] Terzi E, Sarıbacak B, Sağlam F, Cengiz MA. A novel expert system for diagnosis of iron deficiency anemia. Comput Math Methods Med 2022;2022:7352096.

[35] Shander A, Cappellini MD, Goodnough LT. Iron overload and toxicity: the hidden risk of multiple blood transfusions. Vox Sang 2009;97(3):185–97.

[36] Sajewski A. Quantifying Iron Overload using MRI, Active Contours, and Convolutional Neural Networks (Journal Article) | NSF PAGES. 11th Annual Undergraduate Research & Scholarship Symposium. Duquesne University; 2019.

[37] Labranche R, Gilbert G, Cerny M, Vu K-N, Soulières D, Olivié D, et al. Liver Iron quantification with MR imaging: a primer for radiologists. Radiographics. 2018;38 (2):392–412.

[38] Reeder SB, Yokoo T, França M, Hernando D, Alberich-Bayarri Á, Alústiza JM, et al. Quantification of liver Iron overload with MRI: review and guidelines from the ESGAR and SAR. Radiology. 2023 Apr;307(1):e221856.

[39] Huang K-C, Huang C-S, Su M-Y, Hung C-L, Ethan Tu Y-C, Lin L-C, et al. Artificial intelligence aids cardiac image quality assessment for improving precision in strain measurements. JACC Cardiovasc Imaging 2021;14(2):335–45.

[40] Alonso-Fernandez-Gatta M, Martin-Garcia A, Diez-Campelo M, Martin-Garcia AC, Barreiro-Pérez M, Lopez-Cadenas F, et al. Magnetic resonance myocardial feature tracking in transfusion-dependent myelodysplastic syndrome. J Cardiovasc Imaging. 2021;29(4):331–44.

[41] Choi E-Y. Feature tracking analysis, the "cherry-on-top" of cardiac magnetic resonance for suspected Iron overload cardiomyopathy. J Cardiovasc Imaging 2021;29(4):345–6.

[42] Nead KG, Halterman JS, Kaczorowski JM, Auinger P, Weitzman M. Overweight children and adolescents: a risk group for iron deficiency. Pediatrics. 2004;114(1): 104–8.

[43] Stevens GA, Finucane MM, De-Regil LM, Paciorek CJ, Flaxman SR, Branca F, et al. Global, regional, and national trends in hemoglobin concentration and prevalence of total and severe anemia in children and pregnant and non-pregnant women for 1995-2011: a systematic analysis of population-representative data. Lancet Glob Health 2013;1(1):e16–25.

[44] Global anemia reduction efforts among women of reproductive age: impact, achievement of targets and the way forward for optimizing efforts. 2023. Available from: https://www.who.int/publications/i/item/9789240012202.

[46] Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inform 2002;35(5–6): 352–9.

[47] Khan JR, Chowdhury S, Islam H, Raheem E. Machine learning algorithms to predict childhood anemia in Bangladesh. Journal of Data Science 2021;17(1):195–218.

[48] Ford I, McLachlan GJ. Discriminant analysis and statistical pattern recognition. Journal of the Royal Statistical Society Series A (Statistics in Society) 1993;156(2): 327.

[49] Zhao Y, Healy BC, Rotstein D, Guttmann CRG, Bakshi R, Weiner HL, et al. Exploration of machine learning techniques in predicting multiple sclerosis disease course. PloS One 2017;12(4):e0174866.

[50] Wood JC, Enriquez C, Ghugre N, Tyzka JM, Carson S, Nelson MD, et al. MRI R2 and R2* mapping accurately estimates hepatic iron concentration in transfusion-dependent thalassemia and sickle cell disease patients. Blood. 2005;106(4): 1460–5.

[51] Chavhan GB, Babyn PS, Thomas B, Shroff MM, Haacke EM. Principles, techniques, and applications of T2*-based MR imaging and its special applications. Radiographics. 2009;29(5):1433–49.

[52] Littlejohns TJ, Holliday J, Gibson LM, Garratt S, Oesingmann N, Alfaro-Almagro F, et al. The UK biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. Nat Commun 2020;11(1):2624.

[53] Choi JW. Sensitivity, specificity, and predictive value of serum soluble transferrin receptor at different stages of iron deficiency. Ann Clin Lab Sci 2005;35(4):435–9.

Further reading

[45] Hlatswayo BPS, Ntshangase S, De Villiers FPR. The effects of iron deficiency and anemia on primary school learners' scholastic performance. S Afr J CH 2016;10(2): 111.