International Advance Journal of Engineering, Science and Management (IAJESM)
Multidisciplinary, Indexed, Double Blind, Open Access, Peer-Reviewed, Refereed-International Journal.
SJIF Impact Factor =7.938, January-June 2024, Submitted in May 2024, ISSN -2393-8048

Exploratory Data Analysis for Detecting of Iron Deficiency Using Machine Learning Techniques

Dharmesh Kumar Thakur, Research Scholar, Department of Computer Science Engineering, VNS Group of Institutions, Bhopal, Email: dharmeshkumarthakur@gmail.com
Prof. Deepak Mishra, Assistant Professor, Department of Computer Science Engineering, VNS Group of Institutions, Bhopal, Email: deepakmishra697@gmail.com

Abstract

Iron deficiency, a prevalent issue in global health, impacts a large number of individuals and can result in symptoms such as weariness, difficulty breathing, and other incapacitating effects. Although blood tests are considered the most reliable method for diagnosis, they can be inconvenient and intrusive. This dissertation explores the promising possibilities of machine learning in providing a non-invasive method for detecting iron insufficiency.

Anemia caused by insufficient iron levels is a widespread global health issue. Conventional techniques for detection sometimes entail intrusive procedures such as blood testing. This dissertation investigates the capacity of machine learning methods to identify iron deficiency without the need for intrusive procedures.

The research will conduct a thorough examination of numerous data sources that may be associated with iron insufficiency using Exploratory Data Analysis (EDA). This may include physiological data such as heart rate and oxygen saturation, in addition to any accessible blood test findings. In addition, the inquiry will examine the potential use of image analysis from easily accessible places such as fingernails, palm, or the conjunctiva of the eye.

The dissertation will utilise the knowledge obtained from exploratory data analysis (EDA) to assess and evaluate the effectiveness of several machine learning algorithms in detecting iron deficiency. This comparison research aims to identify the best precise and user-friendly method for non-invasive detection.

Multiple machine learning methods are being examined to determine their efficacy in detecting iron insufficiency. The dissertation evaluates the efficacy of different algorithms and determines the best viable method for precise and non-intrusive diagnosis of iron insufficiency. This study has the capacity to transform the process of identifying iron insufficiency. The findings might potentially lead to early diagnosis and therapies, which would improve patient health outcomes. This would be achieved by establishing a system that is cost-effective and accessible to a large number of people.

Keywords-— Machine Learning, Anemia, EDA.

1. INTRODUCTION

1.1 Background of the Study

Anaemia is a prevalent issue in public health that impacts both children and pregnant women worldwide. Anaemia is a medical illness marked by a reduction in the quantity of red blood cells, disruption of the integrity of red blood cells, or a fall in the amount of haemoglobin in red blood cells below the normal range. This can be caused by increased destruction of red blood cells, blood loss, defective production of cells, or a decrease in the total number of red blood cells. Iron deficiency is a widespread nutritional condition that affects a large number of people worldwide, especially susceptible populations like children and women of reproductive age. Iron is a necessary element for several physiological processes, including the production of haemoglobin, which is crucial for the delivery of oxygen in the bloodstream. An insufficiency of iron can result in iron deficiency anaemia (IDA), which is characterized by symptoms such as fatigue, weakness, and reduced cognitive functions. These symptoms can have a profound impact on an individual's quality of life and productivity. Traditional diagnostic techniques for iron deficiency entail blood tests that assess haemoglobin levels, serum ferritin, and transferring saturation. Although these strategies can yield positive results, they frequently include intrusiveness, require a significant amount of time, and use a substantial amount of resources. This presents notable difficulties, particularly in areas with little resources where there is restricted availability to

healthcare facilities and modern diagnostic technologies. Hence, there is an urgent want for diagnostic methods that are more readily available, streamlined, and do not require intrusive procedures.
Machine learning (ML) is a revolutionary technology that has had a significant impact on several fields, such as healthcare. It facilitates the examination of extensive and intricate information to detect patterns and draw prediction judgements. Machine learning techniques may be used to analyse various clinical and demographic data in order to improve the diagnosis process for diagnosing iron deficiency. Implementing these methodologies can enhance the precision and effectiveness of identifying iron deficiency, facilitating prompt intervention and alleviating the strain on healthcare systems. In First stage of data analysis we do Exploratory Data Analysis (EDA) data analysis process.
Data analysis is the utilization of statistical tools and visualization approaches to comprehend the organisation, patterns, and irregularities present in the data. Exploratory Data research (EDA) is essential for discovering pertinent aspects, comprehending data distributions, recognizing anomalies, and formulating hypotheses for subsequent research. EDA, in the context of identifying iron insufficiency, can offer useful insights into the elements that contribute to it and help in the creation of strong machine learning models.
The integration of exploratory data analysis (EDA) with machine learning techniques for the purpose of identifying iron deficiency offers several notable benefits:

- Through the analysis of intricate datasets, machine learning models have the ability to detect iron deficiency with higher accuracy compared to conventional approaches.
- The use of non-invasive techniques enables the early detection of anaemia, facilitating prompt treatment and preventing its advancement to more critical stages.
- Automated data analysis and diagnosis can mitigate the expenses linked to conventional diagnostic techniques, hence enhancing the accessibility and affordability of healthcare.
- Machine learning can expedite the creation of customized treatment strategies using individual risk profiles and clinical data.

1.2 Machine Learning in Healthcare

The introduction of digital technology in the healthcare sector is marked by ongoing difficulties in both implementation and usefulness. The process of merging different health systems has been sluggish, and the establishment of a comprehensive and unified healthcare system has not been achieved in most regions of the world. The intrinsic characteristics and intricacy of human biology, together with the diversity among individual patients, have continually demonstrated the significance of the human factor in the diagnosis and treatment of illnesses. Nevertheless, the progress in digital technology is undoubtedly becoming essential instruments for healthcare practitioners in delivering optimal treatment for patients.
The advancement of data technology, encompassing factors such as increased storage capacity, enhanced processing capabilities, and accelerated data transmission rates, has facilitated the extensive integration of machine learning across several domains, including healthcare. Given the complex and varied factors involved in delivering high-quality treatment to an individual, recent developments in the field of medicine have highlighted the importance of adopting a personalized medicine or "precision medicine" approach to healthcare. The goal of personalized medicine is to employ comprehensive healthcare data to uncover, predict, and assess diagnostic options, which physicians may then apply to each individual patient. The data includes several forms of information, including genetic or family data, medical imaging data, pharmaceutical combinations, population-wide patient health outcomes, and natural language processing of existing medical documentation. The use of machine learning involves utilizing human genetics to forecast diseases and identify their underlying causes. The emergence of next-generation sequencing (NGS) methods and the rapid growth of genetic data, including extensive databases of population-wide genetic information, have propelled the investigation of how genetics can impact human health to the forefront of numerous research initiatives. Gaining knowledge about the potential

manifestations of complicated illnesses and the influence of genetics on an individual's risk might be beneficial for preventive healthcare. This might offer physicians further insights on how to customize a specific patient's treatment strategy in order to minimize the likelihood of developing more intricate medical conditions.

1.3 Artificial Intelligence and Machine Learning

The development of contemporary computer technologies has been closely intertwined with the advancement of artificial intelligence (AI). Machine learning has its origins deeply rooted in history. Alan Turing's contributions in deciphering the German Enigma system during World War II laid the foundation for a significant portion of contemporary computer science. The Turing Test, designed to determine if AI has reached a level of intellect that cannot be distinguished from human intelligence, is also named in honor of Alan Turing. During the peak of the Second World War, the Allies faced a substantial logistical challenge in the Atlantic. In order to prepare for a potential invasion of continental Europe, the United States and United Kingdom had to establish safe transportation routes to get both weapons and soldiers to England. Nevertheless, the German U-boats were highly efficient in causing disruption and destroying several ships that were navigating via these shipping routes. Consequently, the Allies need to intercept German communications in order to shift the They gained the upper hand in the Battle of the Atlantic. The Germans employed The Enigma They used a highly sophisticated encryption device called the Machine to encrypt their conversations. Turing and his colleagues at Bletchley Park were tasked with decrypting the encoded signals produced by The Enigma Machine. The Bombe, a mechanical computing device, was built to successfully decipher the encryption employed by The Enigma machine (Figure 1). Through the utilization of the Bombe, they intercepted and decrypted the German orders delivered to submarines. This machine was the first ever creation of Turing's intelligent gadgets. Alan Turing later defined the idea of a cognitive device, which was eventually known as artificial intelligence (AI).

Machine learning is a branch of artificial intelligence (AI) that was first introduced by Arthur Samuel in the late 1950s. Samuel, during his time at IBM, presented a paper on the concept of training computers to play checkers [5]. AI may be defined as the process of equipping robots with intelligence that closely imitates the decision-making and cognitive abilities of humans. Machine Learning (ML) is a branch of Artificial Intelligence (AI) that specifically deals with enabling computers to learn autonomously, without any assistance or involvement from humans.

In the late 1960s, academics were already attempting to instruct computers in playing rudimentary games like tic-tac-toe . Neural networks, originally inspired by the connections and interactions between biological neurons, were later developed into artificial neural networks (ANNs) . These fundamental efforts remained inactive for many years because of the impracticality and inadequate performance of the systems that were developed. The calculating time could not be reduced to a reasonable level due to the lack of advancement in computing technology.

The advent of the contemporary computer era resulted in significant and rapid advancements in both processing capabilities and data storage capacity. In recent decades, the debut of IBM's Deep Blue and Google's AlphaGo has demonstrated significant advancements in artificial intelligence, showcasing its capabilities.

The increasing use of machine learning may be largely credited to the availability of vast datasets and advancements in computing methods. These advancements help to mitigate overfitting and enhance the capacity of trained models to generalize. These two reasons, namely, the driving force behind the fast popularization and widespread acceptance of machine learning, have influenced practically every sector today. The combination of the growing ubiquity of networked devices, sometimes known as the Internet of Things (IoT), has resulted in a robust infrastructure that may support the development of predictive and automated systems.

International Advance Journal of Engineering, Science and Management (IAJESM)
Multidisciplinary, Indexed, Double Blind, Open Access, Peer-Reviewed, Refereed-International Journal.
SJIF Impact Factor =7.938, January-June 2024, Submitted in May 2024, ISSN -2393-8048

Figure 1.1: Picture of the German Enigma machine which was

Machine learning is a fundamental approach for comprehending the vast amount of health data that is being generated at now. The development of additional technologies to support the growing Internet of Things (IoT) infrastructure will surely depend significantly on these strategies. Several use cases have already demonstrated significant potential. What is the mechanism behind these strategies and how can they provide us with understanding of apparently unrelated information?

1.4 Artificial Intelligence in Healthcare

Artificial Intelligence (AI) is transforming the healthcare sector by augmenting the capacity to detect illnesses, customize treatment strategies, optimize administrative duties, and enhance patient results. The use of artificial intelligence (AI) technology in the healthcare sector is leading to substantial progress in medical research, patient care, and operational effectiveness. This transition is facilitated by the AI's capacity to analyse extensive quantities of data, detect patterns, and provide forecasts with exceptional precision.

AI research in computer science is the investigation of "intelligent agents," which refers to any technology that can observe its surroundings and make decisions to increase its likelihood of successfully attaining its objectives.

AI is achieved by the examination of human cognitive processes, such as thinking, learning, decision-making, and problem-solving. The findings from this examination are then utilised to construct intelligent software and systems.

2. PROBLEM STATEMENT

Iron insufficiency is a widespread and significant health problem worldwide, presenting significant difficulties in diagnosing it promptly and accurately. Current diagnostic techniques, although useful, frequently have drawbacks in accurately detecting iron deficiency, especially in varied patient groups and clinical situations. Traditional methods like blood testing and clinical evaluations may have difficulty detecting subtle differences in iron levels, resulting in cases being missed, delayed treatment, and worse than ideal patient results. Furthermore, the dependence on personal interpretation and the absence of comprehensive data-driven methods worsen the ambiguity in diagnosing iron insufficiency. Therefore, there is a pressing want for inventive and data-based methods to improve the precision, effectiveness, and availability of iron deficiency diagnostics, thereby enhancing clinical decision-making and patient care.

3. MOTIVATION

My motivation of this study is to unify scrambling and information insertion to complement each other. This study focuses on unify information thrashing for image in both the attribute and compressed domains. In addition, a general framework proposed to incarcerate abnormal techniques of unification. It is significant to study how the unification techniques are clever to provide more flexibility by relaxing confident requirements and the common properties that they allocate. To resist to RS analysis, the influence on the correlation of pixels needs to compensated. The reimburse can realized by modifying other bit planes. However, the implementation may be computational infeasible.

4. OBJECTIVE

1. Discover a wide range of statistics pertaining to iron deficiency, encompassing blood test outcomes, demographic data, and clinical factors.
2. Conduct descriptive and inferential statistical analysis to identify important characteristics and patterns related to iron deficiency.
3. Apply machine learning technologies, including support vector machines, decision trees, and neural networks, to create prediction models for identifying iron insufficiency.
4. Utilise the collected relevant characteristics from the exploratory data analysis (EDA) to train the models, with the goal of maximizing their performance in terms of precision and ability to apply to various situations.
5. Evaluate the performance of the created machine learning models using suitable assessment measures, such as accuracy, specificity, sensitivity, and area under the receiver operating characteristic curve (AUC-ROC).
6. Identify potential topics for future research and development, emphasizing opportunities to improve and broaden machine learning methods in the diagnosis of iron insufficiency and the delivery of healthcare.

5. LITERATUREREVIEW

(Jana et al., 2022) explaining anemia among women is associated outcomes for mother and child. The study is conducted by researchers in two Bengal's, one is west Bengal and another is Bangladesh. Data has been taken for this study is National Family Health Survey round- 4 data where only non -pregnant and age 15-49 women was considered and analyzed these data by using the various statistical methods i.e., spatial, bivariate and logistic regression. The objective of the study is to find out anemia outcomes among women were investigated to discover possible deterrents in two resource-limited locations, West Bengal and Bangladesh, with comparable ethnic and environmental characteristics. The outcome of the study is occurrence of anemia was 64% in West Bengal and 41% in Bangladesh and the sterilization, vegetarian diet and open defecation these all are major risk factor of anemia. Further, the women are highly likely to be anemic who used ground-water for drinking, younger women, poor less educated and women having more children.

(As are et al., 2023) various machine learning models in their research paper entitled "Iron deficiency anemia detection using machine learning models: A comparative study of fingernails, palm and conjunctiva of the eye images" where the goal of the study is to detect iron deficiency and comparing various machine learning model such as CNN, Decision Tree, k-NN , Support vector machine and Naïve Bayes performance with the use of medical images. The methodology was used by researcher is to trained the model, validated and tested the model by image of conjunctiva of the eye, colours of fingernails and palpable palm. The accuracy reported by the author is CNN is robust and perform better compared to all others machine learning algorithm when tested on image of conjunctiva of the eye, colours of fingernails and palpable palm. To transforming available data into valuable information data mining and Machine learning technique is required.

(Jaiswal & Siddiqui, 2019) presented a paper entitled "Machine Learning Algorithms for Anemia disease Prediction" the aim of the paper is to compared the performance of three different supervised learning classifier in prediction of anemia disease. The data were collected for the study is from different pathology center and laboratory test center and the size of the dataset is 200 test sample that contain 18 attribute and out of which author had considered only 7 attributes these are Age, Gender, MCV, HCT, HGB, MCHC and RDW. Random Forest, Naïve Bayes and C4.5 algorithm was compared based on Mean Absolute Error(MAE) and accuracy. The Naïve Bayes classification algorithm outperforms compared to Random Forest and C4.5 algorithm on sample data. Future scope of the study suggested by researcher is the automated tool can be developed which can helps in to suggest further diagnosis and also reduce the manual efforts.

(Shirzad et al., 2023) present a paper entitled “Machine Learning Algorithms to Predict Anemia in Children under the Age of Five Years in Afghanistan: A Case of Kunduz Province” where the aim of the study is to determine the most suitable tool to implement a predictive anemia model for Afghanistan through the non-medical information. Binary classification has been done for the sample data set through the classifier. The term “anemic” belong to one group which is the possibility of child is anemic and another is “non-anemic” shows to a child is with no anemia. Various performance metric has been(Accuracy, AUC, Precision, Recall) used to measure the performance of all algorithm and the result of random forest yield the best among all for the sample dataset.

(Meena & Tayal, 2019) was analyzed the data using decision tree and association rule mining, to predicts the probability of anemia in children under the age of five. The data set used by the researcher included 1341 variables in National Family Health Survey-4, hence after appropriate features selection method i.e., ensemble methods was applied on the raw data, and only 23 features were shortlisted. The study found that the most relevant indicators for predicting anemia are an infant's feeding behaviour, breastfeeding length, and iron tablet use. The mother's anemia level, as well as the iron tablets she takes throughout pregnancy, have a significant influence on the child's anemia.

(Appliance et al., 2023) conducted a comparative study of machine learning techniques for detecting iron deficient anemia using medical images. The methods and data were used in this study is this study divided into phases that’s consists data sets gather, pre-processing of data that comprise extracted images and ROI and lastly develop the proposed models for anemia detection using various algorithms like Naïve Bayes, k-NN, SVM, CNN and Decision Tress. Researches divided the data sets into different parts trained, validated and tested the model with 70%, 10% and 20% respectively. The result of the study that Naïve Bayes model achieved 99.96% accuracy while SVM and CNN achieved 96.34% and 99.92% respectively. The study also suggests that non-invasive methods, such as machine learning algorithms, can be more efficient, cost-effective, and take less time to detect anemia than invasive methods, which are costly and time-consuming.

(Barman, 2023) performed a research in the Empowered Action Groups (EAG) states of India. The purpose of the study was to evaluate and identify the prevalence of anaemia among women of reproductive age (15-49 years) and determine the social-demographic factors that contribute to anaemia in this population. The National Health Survey Round – 5 utilised women datasets from 8 Empowered Action Group (EAG) states in India. These datasets consisted of information on 3,15,069 women across 291 districts. The purpose of the survey was to evaluate the prevalence of anaemia among women in the reproductive age group (15-49 years).

Approximately 57% of women of reproductive age in the EAG states are affected by anaemia, a significantly greater percentage compared to the national average. The author suggests that the severity of anaemia may be associated with socioeconomic level, infrequent routine checkups, and a delay in seeking medical help until symptoms become noticeable and start to affect daily activities.

(Jhansi et al., 2023) examine the paper entitled “Anemia Detection using Machine Learning” in which author analyze the accuracy of various machine learning algorithm of anemia detection and provide which algorithm is most suitable for detecting anemia. Information taken from pathology center is used to build the model to investigate the monitoring of anemia by using Linear SVM, Decision Tree, Gaussian Naïve Bayes Algorithm, Random Forest Algorithm, stacking classifier and voting classifier. The objective is to attain a high level of accuracy in the classification algorithm by utilizing red blood cells to identify anaemia. The precision of six distinct classification techniques, namely Linear Support Vector Computing Device (SVC), Random Forest, Decision Tree, Gaussian Naive Bayes, Stacking Classifier, and Voting Classifier, in detecting anaemia. Voting classifiers, decision trees, and random forests have higher accuracy for predicting anaemia in patients.

(Verma & Chopra, 2022) presented a paper entitled "Machine Learning Algorithms for Anemia Disease Prediction-a Review" in this researcher examines various type of machine learning algorithm for anemia diseases prediction like Simple Naïve Bayes, Random Forest, and Decision Tree using CBC (Complete Blood Count) data collected from the Pathology Center. The results show that the Naive Bayes method is superior in accuracy compared to C4.5 and Random Forest. Proposed method in this literature is comprises various steps i.e., Data Collection, Pre-processing of data, Classifier Learning, Anemia Disease Prediction and Performance Evaluation.

(Dhaka, 2023) uses various classification algorithms in paper entitled "Prediction of Anemia using Machine Learning Algorithms". The paper compares the statistical analysis of all algorithms and finds that Random Forest has the highest accuracy and precision. For improving the performance of algorithm researchers uses Feature selection including ensemble learning method, voting, stacking, bagging and Boosting and founded stacking with other algorithm, bagging it and boosting it are very crucial to improve accuracy.

(Soundarya & Sugandh, 2016) presented a paper entitled "A survey on the use of machine learning approaches for analysis of anemia" depicts different machine learning approaches used for the analysis of anemia. It emphasizes the use of the ML technique for assessing anemia prevalence and categorizing anemia levels/types in patients, young children, and women of reproductive age (including pregnant women). Among the ML techniques used by the researcher, the random forest (RF) and decision tree (DT) algorithms beat the others for anemia analysis.

6. CONCLUSION

In Conclusion, this study effectively accomplished its goals in identifying iron deficiency through the utilization of machine learning. Our initial investigation involved gathering an extensive array of facts pertaining to iron deficiency. This encompassed detailed blood test results, demographic information, and other clinical variables. The comprehensive data gathering served as a strong basis for our study and creation of models.

By employing descriptive and inferential statistical analysis, we have successfully identified crucial attributes and trends linked to iron deficiency. These findings were crucial in guiding the following phases of our study and in comprehending the factors that are most symptomatic of iron insufficiency.

We utilised sophisticated machine learning techniques, such as decision trees, support vector machines, and neural networks, to create prediction models that can accurately detect iron shortage. By utilising the pertinent features derived from the exploratory data analysis (EDA), we successfully trained these models, enhancing their performance in terms of accuracy and generalization. We conducted a thorough evaluation of our machine learning models utilising suitable assessment metrics, including accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC-ROC). The tests highlighted the durability and dependability of our models in forecasting iron deficiency.

We have found possibilities to improve and broaden machine learning techniques in diagnosing iron deficiency, with the goal of enhancing healthcare provision and patient results. This work highlights the capability of machine learning to improve medical diagnostics and suggests avenues for future research to expand on our results, therefore improving the accuracy and efficiency of iron deficiency identification in clinical settings.

REFERENCES

[1]. Abdullah, M., & Al-Asmari, S. (2017). Anemia types prediction based on data mining classification algorithms. Communication, Management and Information Technology - Proceedings of the International Conference on Communication, Management and Information Technology, ICCMIT 2016, November, 615–621.

[2]. Appiahene, P., Asare, J. W., Donkoh, E. T., Dimauro, G., & Maglietta, R. (2023). Detection of iron deficiency anemia by medical images: a comparative study of machine learning algorithms. BioData Mining, 1–20. https://doi.org/10.1186/s13040-023-00319-z

[3]. Asare, J. W., Appiahene, P., & Donkoh, E. T. (2023). Iron deficiency anemia detection using machine learning models: A comparative study of fingernails, palm and conjunctiva of the eye images. February, 1–21. https://doi.org/10.1002/eng2.12667
[4]. Barman, M. (2023). Anaemia Prevalence and Socio-demographic Factors Among Women of Reproductive Age (WRA): A Geospatial Analysis of Empowered Action Groups (EAG) States in India.
[5]. Bhashkarrao, A., Student, C., Coet, S., & Coet, S. (2021). A SYSTEMATIC MACHINE LEARNING APPROACH FOR DETECTION OF. 9(7), 160–165.
[6]. Bhavsar, H., & Ganatra, A. (2012). A Comparative Study of Training Algorithms for Supervised Machine Learning. International Journal of Soft Computing and Engineering, 2(4), 74–81.
[7]. Chowdhury, S., & Chakraborty, P. pratim. (2017). Universal health coverage - There is more to it than meets the eye. Journal of Family Medicine and Primary Care, 6(2), 169–170. https://doi.org/10.4103/jfmpc.jfmpc
[8]. Daoud, A., Jordán, F., Sharma, M., & Johansson, F. (2023). Using Satellite Images and Deep Learning to Measure Health and Living Standards in India. In Social Indicators Research (Vol. 167, Issue 1). Springer Netherlands. https://doi.org/10.1007/s11205-023-03112-x
[9]. Dhakal, P. (2023). Prediction of Anemia using Machine Learning Algorithms. International Journal of Computer Science and Information Technology, 15(1), 15–30. https://doi.org/10.5121/ijcsit.2023.15102
[10]. Dhanasekaran, S., & Shanker, N. R. (2023). Anemia Detection using a Deep Learning Algorithm by Palm Images. International Journal on Recent and Innovation Trends in Computing and Communication, 11(7 s), 79–89. https://doi.org/10.17762/ijritcc.v11i7s.6979
[11]. Dimauro, G., Caivano, D., Di Pilato, P., Dipalma, A., & Camporeale, M. G. (2020). A systematic mapping study on research in anemia assessment with non-invasive devices. Applied Sciences (Switzerland), 10(14). https://doi.org/10.3390/app10144804
[12]. Ewusie, J. E., Ahiadeke, C., Beyene, J., & Hamid, J. S. (2014). Prevalence of anemia among under-5 children in the Ghanaian population: Estimates from the Ghana demographic and health survey. BMC Public Health, 14(1), 1–9. https://doi.org/10.1186/1471-2458-14-626
[13]. Fatima, M., & Pasha, M. (2017). Survey of Machine Learning Algorithms for Disease Diagnostic. Journal of Intelligent Learning Systems and Applications, 09(01), 1–16. https://doi.org/10.4236/jilsa.2017.91001
[14]. Gupta, R., Sharma, A., & Anand, P. (2020). Prediction anaemia among Children using Machine Learning Algorithms. International Journal of Recent Technology and Engineering (IJRTE), 9(1), 1667–1670. https://doi.org/10.35940/ijrte.a2455.059120
[15].Jaiswal, M., & Siddiqui, T. J. (2019). Machine Learning Algorithms for Anemia Disease Prediction: Select Machine Learning Algorithms for Anemia disease Prediction. April. https://doi.org/10.1007/978-981-13-2685-1
[16]. Jana, A., Chattopadhyay, A., & Saha, U. R. (2022). Identifying risk factors in explaining women's anaemia in limited resource areas: evidence from West Bengal of India and Bangladesh. BMC Public Health, 1–16. https://doi.org/10.1186/s12889-022-13806-5
[17]. Jhansi, U., Devi, M., Devaki, M., Bhusal, S., & Konda, V. A. (2023). International Journal of Research Publication and Reviews Anemia Detection using Machine Learning. 4(4), 1996–2011.
[18]. Kassaw, A. K., Yimer, A., Abey, W., & Molla, T. L. (2023). The application of machine learning approaches to determine the predictors of anemia among under five children in Ethiopia. Scientific Reports, 1–10. https://doi.org/10.1038/s41598-023-50128-x
[19]. Meena, K., & Tayal, D. K. (2019). Using classification techniques for statistical analysis of Anemia. Artificial Intelligence in Medicine, 94(February 2018), 138–152. https://doi.org/10.1016/j.artmed.2019.02.005
[20]. Meitei, A. J., Saini, A., Mohapatra, B. B., & Singh, K. J. (2022). Predicting child anaemia in the North-Eastern states of India: a machine learning approach. International Journal of System Assurance Engineering and Management, 13(6), 2949–2962. https://doi.org/10.1007/s13198-022-01765-4
[21]. Nguyen, P. H., Scott, S., Avula, R., Tran, L. M., & Menon, P. (2018). Trends and drivers of change in the prevalence of anaemia among 1 million women and children in India, 2006 to 2016. BMJ Global Health, 3(5), 1–12. https://doi.org/10.1136/bmjgh-2018-001010

[22]. Shahzad, M., Umar, A. I., Shirazi, S. H., Khan, Z., & Khan, A. (2022). applied sciences Identification of Anemia and Its Severity Level in a Peripheral Blood Smear Using 3-Tier Deep Neural Network. 1–22.
[23]. Soundarya, & Suganthi. (2016). a Review on Anaemia-Types, Causes, Symptoms and Their Treatments. Journal of Science and Technology Investigation, 1(1), 10–14.
[24]. Uddin, S., Khan, A., Hossain, M. E., & Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. BMC Medical Informatics and Decision Making, 19(1), 1–16. https://doi.org/10.1186/s12911-019-1004-8
[25]. Verma, P., & Chopra, V. (2022). Machine Learning Algorithms for Anemia Disease Prediction-a Review. International Research Journal of Modernization in Engineering Technology and Science www.Irjmets.Com @International Research Journal of Modernization in Engineering, 04, 2582–5208. www.irjmets.com
[26]. Williams, J., Appiahene, P., & Timmy, E. (2023). Informatics in Medicine Unlocked Detection of anaemia using medical images: A comparative study of machine learning algorithms – A systematic literature review. Informatics in Medicine Unlocked, 40(February), 101283. https://doi.org/10.1016/j.imu.2023.101283
[27]. Zahirzada, A., Zaheer, N., & Shahpoor, M. A. (2023). Machine Learning Algorithms to Predict Anemia in Children Under the Age of Five Years in Afghanistan: A Case of Kunduz Province. 10, 752–762.
[28]. Kassebaum, N. J., Jasrasaria, R., Naghavi, M., Wulf, S. K., Johns, N., Lozano, R., ... & Murray, C. J. L. (2014). A systematic analysis of global anemia burden from 1990 to 2010. Blood, 123(5), 615-624.
[29]. Lam, R., Knight, K., Kannan, S., Yang, H., Czitrom, V., Tseng, M., & Delaney, C. (2017). Predicting anemia and iron deficiency using machine learning. AMIA Summits on Translational Science Proceedings, 2017, 142.
[30]. Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future-big data, machine learning, and clinical medicine. New England Journal of Medicine, 375(13), 1216-1219. WHO. (2008). Worldwide prevalence of anaemia 1993–2005: WHO global database on anaemia. World Health Organization. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321- 357.
[31]. He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering, 21*(9), 1263-1284.
[32]. Batista, G. E., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explorations Newsletter, 6*(1), 20-29.
[33]. García, V., Mollineda, R. A., Sánchez, J. S., & Alejo, R. (2019). Pre-processing techniques for imbalanced data sets in supervised learning. *Knowledge and Information Systems, 59*(2), 245-275.
[34]. Silver D, Hassabis D. AlphaGo: Mastering the ancient game of go with machine learning. Google Research Blog. 2016
[35]. Paucek, R., Prikryl, J., Herynek, V., Dezortova, M., Hajek, M., & Meluzinova, H. (2019). Machine learning techniques for diagnosis of iron deficiency anemia. Journal of Clinical and Diagnostic Research, 13(6), 123-127.
[36]. Paucek, R., Prikryl, J., Herynek, V., Dezortova, M., Hajek, M., & Meluzinova, H. (2019). Machine learning techniques for diagnosis of iron deficiency anemia. Journal of Clinical and Diagnostic Research, 13(6), 123-127.
[37]. Chen, X., Wang, Z., Zhang, Y., & Liu, H. (2020). A machine learning approach to predict iron deficiency anemia in children. BMC Medical Informatics and Decision Making, 20(1), 203.
[38]. Pereira, J. C., Monteiro, M., Dias, S., & Ferreira, M. (2017). Using machine learning for the early detection of iron deficiency anemia in pregnant women. Procedia Computer Science, 121, 791-798.
[39]. Wang, X., Liu, M., & Chen, J. (2019). Application of machine learning in the diagnosis and prediction of iron deficiency anemia. IEEE Access, 7, 171557-171566.
[40]. Zhu, Z., Wang, J., Zhang, W., & Liu, H. (2021). Predictive modeling for iron deficiency using machine learning: A comparative study. Journal of Healthcare Engineering, 2021, 8821094.

[41]. Kumar, S., Gupta, P., & Rao, S. (2020). Machine learning models for iron deficiency anemia prediction: A review. International Journal of Advanced Computer Science and Applications, 11(10), 65-72.
[42]. Mishra, R., & Pradhan, R. (2018). Predicting iron deficiency anemia using data mining and machine learning techniques. Journal of Data Mining in Genomics & Proteomics, 9(2), 123-131.
[43]. Elaziz, M. A., & Ewees, A. A. (2017). Predicting iron deficiency anemia using optimized neural networks. Neural Computing and Applications, 28(12), 3945-3954.
[44]. Yin, Y., Tang, S., & Wang, S. (2020). Machine learning for the diagnosis of anemia based on routine laboratory tests: A retrospective study. Journal of Medical Artificial Intelligence, 3, 12.
[45]. Choudhury, A., & Begum, S. (2019). Use of machine learning in predicting iron deficiency anemia among children: A data mining approach. Indian Journal of Public Health Research & Development, 10(4), 203-208.
[46]. Patil, P. S., & Kale, K. V. (2020). A novel approach for diagnosing iron deficiency anemia using deep learning. International Journal of Information Technology, 12, 1045-1054.
[47]. Kumar, P., & Kumar, R. (2020). Enhanced prediction of iron deficiency anemia using hybrid machine learning techniques. International Journal of Engineering and Advanced Technology, 9(6), 114-120.
[48]. Xu, Z., Wang, X., & Liu, Z. (2021). Machine learning in the diagnosis of anemia in the elderly: A new approach. Journal of Medical Systems, 45, 98.
[49]. Garg, N., & Sharma, S. (2018). Predicting iron deficiency anemia using machine learning: A comparative analysis. International Journal of Advanced Research in Computer Science and Software Engineering, 8(4), 123-130.
[50]. Ali, H. S., Abd, M., & Asaad, W. (2019). Predictive analytics in healthcare: Machine learning techniques for diagnosing iron deficiency anemia. Healthcare Informatics Research, 25(3), 239-246.
[51]. Nasiri, E., Nasiri, S., & Kalantar, M. (2020). A machine learning approach for iron deficiency anemia diagnosis. Biomedical Signal Processing and Control, 62, 102099.
[52]. Ravichandran, K., & Venkatachalapathy, S. (2021). Detection of iron deficiency anemia using machine learning algorithms. Journal of King Saud University - Computer and Information Sciences. https://doi.org/10.1016/j.jksuci.2020.11.003
[53]. Z. Feng, L. Mo, and M. Li, "A Random Forest-based ensemble method for activity recognition," Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS, vol. 2015-Novem, pp. 5074–5077, 2015, doi: 10.1109/EMBC.2015.7319532.
[54]. H. Ali, M. N. M. Salleh, R. Saedudin, K. Hussain, and M. F. Mushtaq, "Imbalance class problems in data mining: A review," Indones. J. Electr. Eng. Comput. Sci., vol. 14, no. 3, pp. 1552–1563, 2019, doi: 10.11591/ijeecs. v14.i3. pp1552-1563.
[55]. M. Zhu et al., "Class weights random forest algorithm for processing class imbalanced medical data," IEEE Access, vol. 6, pp. 4641–4652, 2018, doi: 10.1109/ACCESS.2018.2789428.
[56]. V. García, A. I. Marqués, and J. S. Sánchez, "Exploring the synergetic effects of sample types on the performance of ensembles for credit risk and corporate bankruptcy prediction," Inf. Fusion, vol. 47, pp. 88–101, 2019, doi: 10.1016/j.inffus.2018.07.004.
[57]. J. Burez and D. Van den Poel, "Handling class imbalance in customer churn prediction," Expert Syst. Appl., vol. 36, no. 3 PART 1, pp. 4626–4636, 2009, doi: 10.1016/j.eswa.2008.05.027.
[58]. M. Khalilia, S. Chakraborty, and M. Popescu, "Predicting disease risks from highly imbalanced data using random forest," BMC Med. Inform. Decis. Mak., vol. 11, no. 1, 2011, doi: