Remittances Review
September 2024,
Volume: 9, No: S 4, pp. 605-616
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

Received : 20 July 2024, Accepted: 25 September 2024
DOI: https://doi.org/10.33282/rr.vx9i2.34

Predictive Analysis of Thalassemia Risk using Statistical and Machine Learning Approaches

[1]Muhammad Azhar Mushtaq, [2]Faisal Bukhari, [3]Sadaqat Ali Ramay, [1]Muddassar Ali Zaidi, [3]Tahir Abbas Khan, [3]Sayyid Kamran Hussain, [4]Faheem Ahmed Khalid, [2]Ahmad Makki, [2]Abu Huraira, [1]Muhammad Kaleem

1st Department of Information Technology, Faculty of Computing, University of Sargodha, Sargodha, 40100, Pakistan.
2nd Department of Data Science Faculty of Computing and Information Technology University of the Punjab
Lahore, Pakistan
3rd Department of Computer Science, TIMES Institute, Multan, Pakistan.
4th Department of Computer Science & I. T Thal University Bhakkar

Abstract*—Thalassemia is a hereditary condition where the body is unable to manufacture enough hemoglobin. Made up of alpha and beta globin proteins, hemoglobin is the most important component of red blood cells (RCB) that delivers oxygen throughout the body. Alpha and beta-globin genes are either rare or nonexistent, which results in alpha and beta- thalassemia. Beta thalassemia is more dangerous because of the increase in the probability of conceiving a kid with thalassemia than the alpha one. Most forms of thalassemia cause chronic and lifelong anemia that exists in early childhood and requires a blood transfusion due to deformity of blood cells frequently throughout the patient's life. The body makes glucose as a result of the oxygen carried by red blood cells, which enables normal body function. Thus, thalassemia impacts the body's ability to distribute oxygen to all of its cells, which can have an impact on organs with severity and even cause death. According to the research anemia caused affects 42% of women worldwide, including 52% of pregnant women in developing nations, compared to 23% in developed economies. In this study, machine learning and statistical analysis are used to forecast and assess the behavior of thalassemia. Moreover, the person with thalassemia should be referred to proper genetic counseling. The person with the alpha thalassemia trait has a normal life expectancy. People with beta-thalassemia often die by the age of 30. The statistical analysis applied in our research are the Independent Samples t-test for Age, the Paired Samples t-test for Hemoglobin (HGB) Levels, Analysis of Variance (ANOVA) for Mean Corpuscular Volume (MCV) Levels Across Age Groups, and the Comparison of Two Hypotheses with Different Means. Moreover, we also investigate the correlation between Red Blood Cells and Hemoglobin. As for the machine learning approaches, we applied supervised machine learning models, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN).

***Index Terms*—Thalassemia, machine learning, statistical modeling, predictive analysis.**

Remittances Review
September 2024,
Volume: 9, No: S 4, pp. 605-616
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

I. INTRODUCTION

Thalassemia is a genetic condition in which the body produces faulty hemoglobin, which destroys red blood cells and causes anemia. Hemoglobin consists of two proteins: four alpha-globin and two beta-globin proteins. There are two types of hemoglobin, alpha and beta hemoglobin. Alpha- thalassemia is less hazardous than beta-thalassemia. Beta thalassemia is more effective for the body because of the major effect and has more probability of conceiving a kid with thalassemia minor or major. [4], [13], [18]

Thalassemia causes chronic and lifelong anemia that can begin in early childhood and often requires frequent blood transfusions due to the deformity of red blood cells through- out the patient's life unless he or she can receive a 100% matched bone marrow transplant, which is unlikely. Thus, thalassemia treatment consumes the most blood compared to other treatments that require blood frequently. Anemia is expected to affect 42% of all women worldwide, including 52% of pregnant women in impoverished nations, compared to 23% in wealthy economies. It has been estimated that Anemia accounts for 3.7% and 12.8% of maternal mortality during pregnancy and childbirth in Africa and Asia, respectively, and has been linked to low neonatal birth weight. [10], [16] Thalassemia has three categories major, minor, and mild or Intermediate. In major thalassemia, individuals require regular and lifelong blood transfusions. Due to the severity of anemia in thalassemia major patients, it is necessary to transfuse blood to maintain an adequate supply of desired hemoglobin to the patient. These transfusions are usually started in the early stages of childhood and are performed regularly, usually after every 2 to 4 weeks. In mild thalassemia patients, the need for transfusions varies from patient to patient. [5], [22] Some patients with thalassemia intermedia may not need to transfuse blood regularly and can manage the bear the condition without them. Others may need transfusions, especially during illness or increased stress. In thalassemia minor patients do not require any blood transfusions and maybe they do not even have mild or any symptoms, and they can lead normal, healthy lives without transfusion of blood. Regular blood transfusions are a popular treatment for thalassemia. Again, bone marrow transplantation is an essential therapy for thalassemia. Now in the modern era of artificial intelligence and machine learning, everyone understands the importance of using it in the medical field as well. The ML algorithm takes the old data as input and examines the new output value. As a result, this thalassemia is predicted using machine learning. Patients with severe thalassemia die early. Major thalassemia has resulted in a variety of consequences. According to research, most thalassemia patients can survive for 25 to 30 years if they are supplied with enough conditions to live until the age of
60. [6], [21]

The primary areas of machine learning applications include image or speech recognition, medical diagnosis, statistical difficulties, predictive analytics, and extraction. ML is exciting and humorous because programs learn from examples. On the other hand statistical analysis is helpful in comprehensively analyzing Thalassemia, enabling us to identify the significant risk factors like genetic and demographic variables. It also enables us to facilitate correlation studies on relationships between health metrics and Thalassemia reports. Contribute to diagnostic assessments and predictive modeling for risk prediction of Thalassemia. [9], [19], [20]

Furthermore, statistical methods explain the population studies, compare results across different groups, and aid in treatment outcomes. The burden of thalassemia on healthcare resources may be quantified using statistical estimates of blood transfusion requirements. The significant influence of anemia on maternal mortality rates, as revealed by statistical analysis, emphasizes the critical need for predictive models to identify at-risk individuals.

I. PREVIOUS WORK

In 2009, in an article by Muncie and Campbell, alpha and beta thalassemia were explored as hereditary hematologic disorders causing hemolytic anemia due to imbalances in globin chain synthesis. The authors emphasize the need for family physicians to diagnose and differentiate thalassemia from other causes of microcytic anemia. Epidemiological insights reveal varying prevalence across ethnic groups, with alpha or beta thalassemia trait affecting 1.7% globally. The article explains into the psyche of hemoglobin composition and transitions, detailing manifestations of alpha thalassemia (silent carrier, trait, HbH disease) and beta-thalassemia (trait, major,

intermedia). Clinical recommendations include cautious iron supplement use and blood transfusion therapy. While providing valuable insights, the article's limitations lie in its clinical focus, suggesting potential for further research incorporating genetic and molecular perspectives to enhance our understanding of thalassemia. [15], [17]

In 2010, The study by Nishi Madan and the team looked at how common a certain blood condition (Beta-thalassemia trait) is in different parts of northern and western India, especially among schoolchildren. India has a diverse population, and this study wanted to understand the frequency of this blood condition in various regions. They found that in Mumbai and Delhi, about 4 out of every 100 schoolchildren have this condition. The study showed that the prevalence varies between regions and different groups of people. The researchers also checked for anemia, a condition where the blood doesn't have enough red cells, and found differences between boys and girls in Mumbai and Delhi. The study suggests that it's important to have programs to check for this condition, especially in regions where it's more common. So this study was only focused on a specific region and all aspects were applied using different machine learning models. [3], [14]

In 2011, Galanello and Cao delve explored the pattern of alpha-thalassemia, a genetic abnormality characterized by reduced or absent production of alpha globin chains in blood. The research of alpha-thalassemia purpose was to explore it at both clinical and molecular levels. They highlight its prevalence in regions historically afflicted by malaria-like diseases and explain clinical challenges posed by population migrations into new regions. Severe diseases like HbH disease and Hb Bart hydrops fetalis are miraculously explored. The authors underscore the importance of genotype- phenotype correlation and provide insights into the diverse field across different alpha-globin genotypes. Molecular genetics, particularly deletions involving alpha globin genes, is dissected, revealing the spectrum of defects contributing to alpha-thalassemia. This review acts as a valuable resource for researchers to navigate the complex behavior of alpha thalassemia. [1], [8]

In 2016, The study by Khair-allah, Balkis A., and the team looked at the connection between a person's genes (genotype) and the severity factors of Beta Thalassemia Major (TM), a disorder related to issues with globin genes. We used a method called decision-making trial and evaluation laboratory (DEMATEL) to explore how different factors, like mutations, are related to the severity of TM. The goal is to understand these relationships and use the information to improve medical diagnosis and figure out the best treatment. We created graphs to show the connections between these factors. Keywords: bioinformatics, Beta Thalassemia major, genotype, correlation, DEMATEL. [11], [12] In 2019, The study by Egejuru, Ngozi, and the team used data mining algorithms to predict the risk of thalassemia in people of all ages. They gathered information on thalassemia risk factors through interviews and questionnaires with medical experts. The data collected included demographic and clinical variables. Supervised machine learning algorithms were used to create a predictive model for thalassemia risk, simulated using the Waikato Environment for Knowledge Analysis (WEKA). The model was validated with historical hospital data. The study found that the risk factors included gender, age, marital status, ethnicity, family history, spleen enlargement, diabetes, urine color changes, and parent carriers. The distribution of thalassemia risk was categorized as no cases (43%), low cases (10%), moderate cases (16%), and high cases (31%). The study concluded that using the multi-layer perceptron for thalassemia prediction could enhance healthcare decision-making. Keywords: Thalassemia, Anemia, Predictive Model, Naïve Bayes, Classifier, Multilayer Perceptron. [2], [7]

II. MATERIALS AND METHODS

A. Study Area

The research employed a cross-sectional, quantitative de- sign, focusing on an adult population selected from various areas in Pakistan, particularly Karachi and rural areas. The target group comprised youth aged 1-30. The study involved
129 respondents, with an intentional over-representation of individuals aged 1-44. Among them, 93 were male, and 36 were female, indicating a higher ratio of male participants in the survey. The dataset includes observations of both normal patient behavior and those affected.

B. Data Collection

The data collection process for this study incorporated both primary quantitative and qualitative data. We gathered information through collaboration with the Royal Blood Society and Shoukat Khanam Hospital. To maintain privacy, face-to- face interactions with patients and doctors were conducted after obtaining consent. The research focuses on various parameters such as gender, age, RBC, HGB, HCT, MCV, MCH, MCHC, and Diagnosis.

C. Machine Learning Models

In our research, we used special computer programs to learn from data and make predictions. These programs, like Random Forest, Support Vector Machine (SVM), and K- Nearest Neighbors (KNN), help us understand information better and improve our ability to foresee outcomes.

D. Statistical Analysis

For the Statistical Analysis in our research, we use the Python language and perform tests in the Anaconda Notebook. The conducted statistical tests encompass the Independent Samples t-test for Age, the Paired Samples t-test addressing Hemoglobin (HGB) Levels, Analysis of Variance (ANOVA) scrutinizing Mean Corpuscular Volume (MCV) Levels across various Age Groups, and a Comparative Analysis of Two Hypotheses with distinct Means. Furthermore, our exploration extends to probing the correlation existing between Red Blood Cells and Hemoglobin.

III. MACHINE LEARNING APPROACH

A. Information of Participants

The details of the participants are outlined in Table I, where the total number of participants was 129. Among them, 93 (72%) were male, and 36 (28%) were female. In terms of age distribution, 28 participants (22%) fell in the 1-18 age category, 57 participants (44%) were in the 18-30 age range, and 39 participants (34%) were in the 30-50 age group.

TABLE I: Information of Participants

Variable	Participants=129	Percentage (%)
Gender		
Male	93	72%
Female	36	28%
Age		
1-18	28	22%
18-30	57	44%
30-50	39	34%

B. Significance in Research

In machine learning approach, we focused on examining the details of participants, considering various parameters to observe their behavior and significance. By employing a predictive approach, we aimed to determine the age category of patients, distinguish between normal and affected individuals, identify the most influential parameter in thalassemia, and understand the least important one. Additionally, we explored whether males or females are more prone to this disease.

C. Practical Implementation

After carefully examining the data in our research, Machine learning models, including Random Forest, Logistic Regression, K-Nearest Neighbors, and Support Vector Machine were applied to observe the behavior. These models were chosen to predict and analyze outcomes based on the collected information. The results, depicting the accuracy of each model, are visually presented in Figure II. This practical implementation allowed us to assess the

Remittances Review
September 2024,
Volume: 9, No: S 4, pp. 605-616
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

effectiveness of different machine learning techniques in handling the data and making predictions related to our research objectives.

D. Research Results

Our research results highlight which factors or parameters are most important in predicting thalassemia. Figure I shows the features that are highly likely to indicate thalassemia, helping us understand which ones matter the most. Figure II presents the ages of people affected by the disease, revealing which age groups are more prone to thalassemia. Our model, once implemented, accurately predicted thalassemia with a high 97% success rate. Figure III provides details on the accuracy of different models, showing their effectiveness in predicting thalassemia.

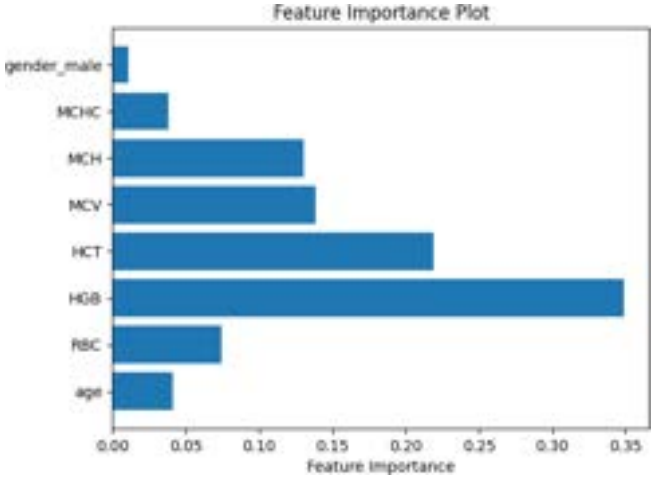


Fig. 1: Feature Importance Plot

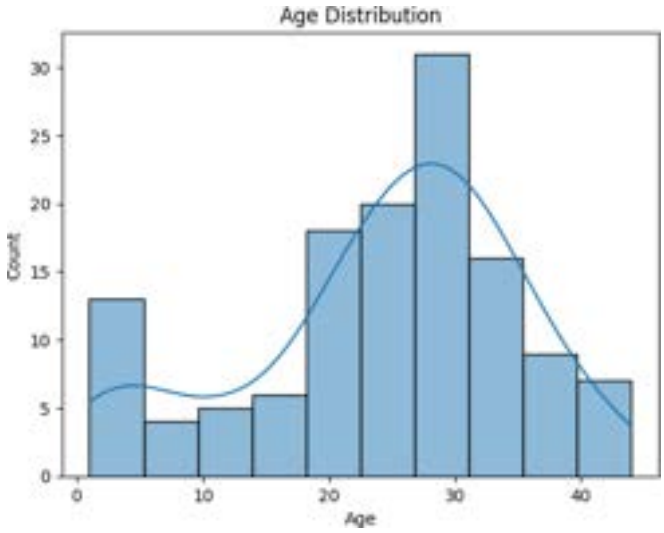
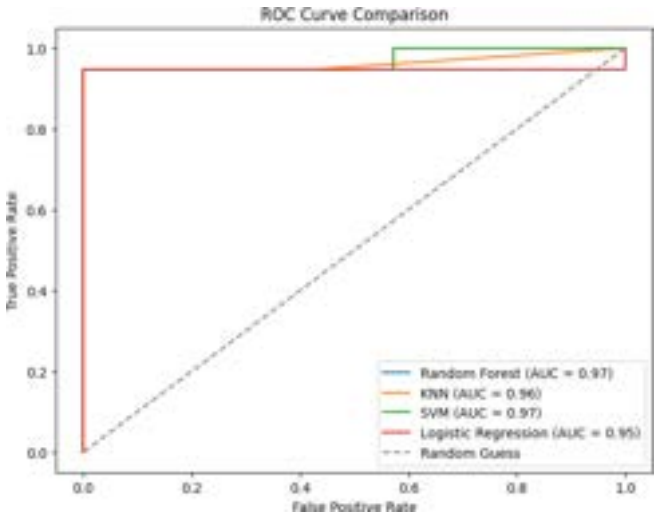


Fig. 2: Results in Age Distribution

high 97% success rate. Figure III provides details on the accuracy of different models, showing their effectiveness in predicting thalassemia.

The performance of four machine learning classifiers, namely Random Forest, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Logistic Regression, was evaluated in the context of thalassemia prediction. The Random Forest classifier demonstrated an overall accuracy of 92%, achieving a balanced precision and

recall for both positive and negative classes. KNN exhibited superior performance with an accuracy of 96%, particularly excelling in correctly classifying individuals without thalassemia. SVM, while achieving an accuracy of 88%, demonstrated slightly lower precision for the positive class. Logistic Regression mirrored KNN's accuracy of 96%, showcasing robust predictive capabilities. The classification reports provide a comprehensive overview of each classifier's precision, recall, F1-score, and support metrics, aiding in the assessment of their efficacy in thalassemia prediction.



In summary, the machine learning models, especially KNN and Logistic Regression, displayed promising results in accurately predicting thalassemia cases. These findings suggest the potential utility of these models in clinical settings for early detection and intervention. Further analysis, including the consideration of additional features and larger datasets, could contribute to refining the predictive capabilities of these models for enhanced diagnostic accuracy in thalassemia screening.

E. Impact of Machine Learning in Our Approach

Incorporating machine learning into our research approach has significantly impacted the accuracy and efficiency of thalassemia prediction. The utilization of supervised machine learning algorithms, such as Random Forest, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Logistic Regression, enabled the creation of predictive models based on essential parameters. The models demonstrated a remarkable ability to discern patterns within the dataset, providing valuable insights into the probability of thalassemia across different age groups and genders. The predictive accuracy of these models, particularly exemplified by KNN and Logistic Regression with 96% accuracy, suggests their potential application in real-world scenarios for early identification of thalassemia cases.

Furthermore, the impact of machine learning in our research extends beyond predictive accuracy to encompass

the interpretability of results. The classification reports generated by each model offer a detailed breakdown of precision, recall, and F1-score for both positive and negative classes.

This transparency allows clinicians and researchers to understand the strengths and limitations of each model, facilitating informed decision-making in a clinical context. The integration of ma- chine learning not only enhances the precision of thalassemia prediction but also empowers healthcare professionals with valuable information to tailor appropriate interventions and treatments based on the individualized risk profiles predicted by these models.

TABLE II: Random Forest Classifier Results

	Precision	Recall	F1-Score	Support
0	0.78	1.00	0.88	7
1	1.00	0.89	0.94	19
Accuracy		0.92		
Macro Avg	0.89	0.95	0.91	26
Weighted Avg	0.94	0.92	0.93	26

TABLE III: KNN Classifier Results

	Precision	Recall	F1-Score	Support
0	0.88	1.00	0.93	7
1	1.00	0.95	0.97	19
Accuracy		0.96		
Macro Avg	0.94	0.97	0.95	26
Weighted Avg	0.97	0.96	0.96	26

TABLE IV: SVM Classifier Results

	Precision	Recall	F1-Score	Support
0	0.70	1.00	0.82	7
1	1.00	0.84	0.91	19
Accuracy		0.88		
Macro Avg	0.85	0.92	0.87	26
Weighted Avg	0.92	0.88	0.89	26

TABLE V: Logistic Regression Classifier Results

	Precision	Recall	F1-Score	Support
0	0.88	1.00	0.93	7
1	1.00	0.95	0.97	19
Accuracy		0.91		
Macro Avg	0.94	0.97	0.95	26
Weighted Avg	0.97	0.96	0.96	26

IV. STATISTICAL ANALYSIS

A. Chi-Square Test of Association (Gender and Diagnose)

The Chi-Square test was conducted to examine the association between gender and the diagnosis status (patient or normal). The obtained p-value of 0.0001 is less than the significance level of 0.05, leading to the rejection of the null hypothesis. This indicates a statistically significant association between gender and the likelihood of being diagnosed as a patient.

B. Kruskal-Wallis Test (HGB Levels and Diagnose)

The Kruskal-Wallis test was performed to assess if there is a significant difference in hemoglobin (HGB) levels among different diagnosis groups. The extremely low p-value (3.8e- 15) suggests rejecting the null hypothesis, indicating that there are statistically significant differences in HGB levels between patients and normal groups.

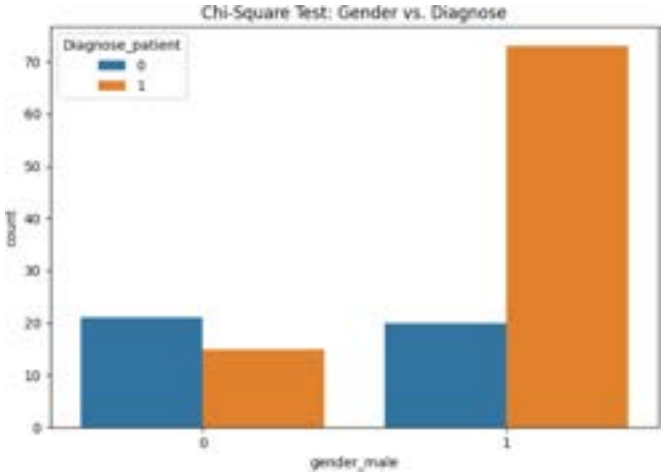


Fig. 4: Chi-Square Test

C. Independent Samples t-test (Age and Diagnose)

An independent samples t-test was utilized to investigate whether there is a significant difference in age between patients and normal individuals. With a p-value of 0.0006, below the significance level, the null hypothesis is rejected. This implies that there is a statistically significant age difference between the two groups.

D. Paired Samples t-test (HGB Levels)

The paired samples t-test was applied to evaluate if there was a significant difference in HGB levels within the same individuals. However, the test resulted in NaN, indicating insufficient data or variability in the HGB levels for this specific analysis.

E. Analysis of Variance (ANOVA) (MCV Levels and Age Groups)

ANOVA was employed to examine the potential impact of age groups on mean corpuscular volume (MCV) levels. However, the test resulted in NaN, suggesting that there might be insufficient data or variability in MCV levels across different age groups.

F. Correlation Analysis (RBC and HGB)

Correlation analysis was performed to assess the relation- ship between red blood cell count (RBC) and hemoglobin (HGB) levels. The correlation coefficient of -0.11979 and a p-value of 0.1763 indicate a weak, non-significant negative correlation between RBC and HGB levels. Therefore, the null hypothesis of no correlation is not rejected.

V. DISCUSSION

The discussion of our research findings underscores the impactful role of machine learning models, specifically Random Forest, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Logistic Regression, in predicting thalassemia. The consistently high accuracy across these models, reaching up to 96%, signifies their effectiveness in discerning patterns within the dataset and identifying individuals at risk. This predictive accuracy, coupled with detailed classification reports, enhances the interpretability of the results, providing valuable insights into the performance of each model.

Remittances Review
September 2024,
Volume: 9, No: S 4, pp. 605-616
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

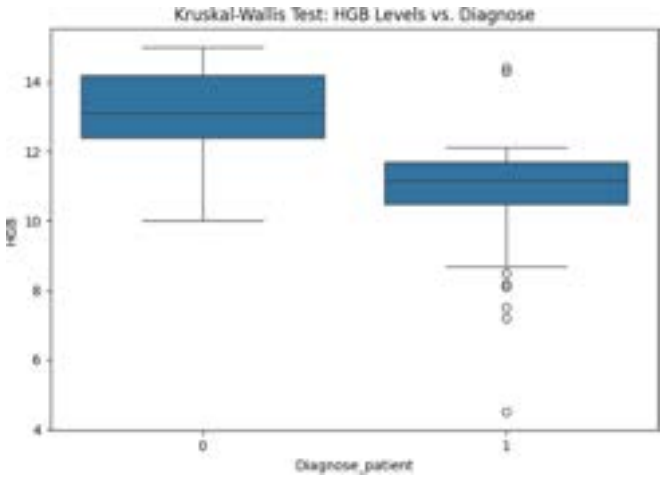


Fig. 5: Kruskal-Wallis Test

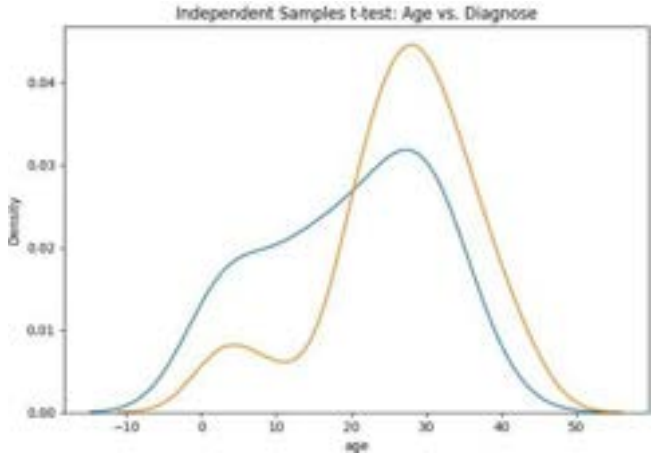


Fig. 6: Independent Samples t-test

The utilization of machine learning approaches allowed us to evaluate the significance of various parameters, including age and gender, in predicting thalassemia. The comprehensive analysis revealed that these models not only accurately predict thalassemia but also shed light on the influence of different factors on the disease's manifestation. For instance, the age distribution analysis showcased the prevalence of thalassemia across various age groups, offering a nuanced understanding of the disease's impact within the studied population.

In the context of clinical applications, the robust predictive capabilities of machine learning models offer promising avenues for early thalassemia identification. The ability to tailor interventions based on individualized risk profiles derived from these models can significantly enhance patient outcomes. However, it is crucial to

acknowledge the need for further validation and integration of these models into clinical practice. Collaborative efforts between data scientists, healthcare professionals, and regulatory bodies are essential to ensure the ethical and responsible deployment of machine learning technologies in the realm of thalassemia diagnosis and management.

In conclusion, our research demonstrates the transformative potential of machine learning in thalassemia prediction, pro- viding accurate and interpretable models that can contribute to more effective healthcare strategies. The findings open avenues for future research to refine and expand the application of machine learning in the broader context of genetic disorders, ultimately improving patient care and outcomes.

The statistical analysis conducted in our research provides valuable insights into the relationships between various parameters and the diagnosis of thalassemia. The Chi-Square test revealed a significant association between gender and the likelihood of being diagnosed as a patient. This emphasizes the importance of considering gender-specific factors in thalassemia diagnosis. Additionally, the Kruskal-Wallis test demonstrated significant differences in hemoglobin (HGB) levels between patients and normal groups, highlighting HGB as a crucial diagnostic indicator.

The Independent Samples t-test unveiled a statistically significant age difference between patients and normal individuals, suggesting that age is a relevant factor in thalassemia susceptibility. However, the paired samples t-test for HGB levels resulted in NaN, indicating a need for further investigation and potential data refinement in this specific analysis. Similarly, ANOVA for mean corpuscular volume (MCV) levels across different age groups yielded NaN, prompting a closer examination of data variability.

In correlation analysis, the weak, non-significant negative correlation between red blood cell count (RBC) and HGB levels suggests that these parameters may provide complementary information rather than exhibit a direct linear relationship. While statistical tests have elucidated certain relationships, the presence of NaN values underscores the importance of meticulous data handling and potential limitations in the dataset. Future studies should focus on refining data collection methods and addressing these limitations to enhance the robustness of statistical analyses in the context of thalassemia diagnosis.

I. Limitations

This research significantly contributes to the understanding of thalassemia by analyzing key parameters including Hematocrit (HCT), Mean Corpuscular Volume (MCV), Mean Corpuscular Hemoglobin (MCH), and Mean Corpuscular Hemoglobin Concentration (MCHC), it is crucial to acknowledge certain limitations. The reliability of our findings is possible on the quality and availability of the data, and incomplete or inaccurate datasets may introduce biases into our analysis. The sample size and representatives of our study population could influence the external validity of our results. Also, the application of machine learning models trains inherent challenges such as over-fitting, sensitivity, and the necessity for large datasets. Our focus on specific parameters may not contain the full clinical data of thalassemia, and there might be possible, that there could be another parameter that could play a role in thalassemia and missed. Despite these limitations, our research provides valuable insights into thalassemia, and future studies should address these constraints to advance our understanding and enhance the clinical utility of our findings.

I. Conclusion

In conclusion, our research investigates the complicated aspects of thalassemia, a hereditary condition that significantly impacts the body's ability to produce sufficient hemoglobin. Our research employed a different approach, combining tra- ditional statistical analysis and different machine-learning techniques to reveal patterns and behaviors which is re- lated to thalassemia. Statistical tools such as the Independent Samples t-test, Paired Samples t-test, Analysis of Variance (ANOVA), and Comparison of Two Hypotheses with Different Means provided valuable insights into age-related differences, hemoglobin levels, and mean corpuscular volume across var- ious age groups. Additionally, the correlation between Red Blood Cells and Hemoglobin sheds light on the intricate dynamics within thalassemia-affected individuals. Also, the results show the parameters that mainly play a role in tha- lassemia are Hematocrit (HCT), Mean Corpuscular Volume (MCV), Mean Corpuscular Hemoglobin (MCH), Mean Corpuscular Hemoglobin Concentration (MCHC). These parame- ters are crucial in detecting the type of anemia and

thalassemia in patients. Abnormal values in these measures can lead to detect the exact blood disorder in a patient. In addition, our research relates the traditional and mod- ern analytical approaches to comprehensively understand and address the complexities of thalassemia. The findings hold implications for both clinical interventions and public health strategies, exploring the way for more targeted and effective management of this hereditary condition.

REFERENCES

[1] S Ahmed, M Saleem, B Modell, and M Petrou. Screening extended families for genetic hemoglobin disorders in pakistan. *N Engl J Med*, 347:1162–1168, 2002.
[2] S. R. Amendolia et al. A comparative study of k-nearest neighbour, support vector machine and multi-layer perceptron for thalassemia screening. *Chemometrics and Intelligent Laboratory Systems*, 69:13– 20, 2003.
[3] AH Bittles. The role and significance of consanguinity as a demographic variable. *Popul Dev Rev*, 20:561–584, 1994.
[4] A. Cao and R. Galanello. Beta-thalassemia, genetics in medicine ". *Computer methods and programs in biomedicine*, 12:61–76, Oct 2009.
[5] Gallanello R. Cao A, Rasatelli MC. Beta thalassemia by carrier screening genetic counseling and prenatal diagnosis. pages 137–197, 2007.
[6] M. D. Capellini, A. Taher, et al. *Guidelines for the management for transfusion of Thalassemia dependent (TDT)*. Thalassemia International Federation (TIF), 3rd edition, 2014.
[7] Ngozi Chidozie Egejuru, Sekoni Olayinka Olusanya, Adanze Onyenonachi Asinobi, Omotayo Joseph Adeyemi, Victor Oluwatimilehin Adebayo, and Peter Adebayo Idowu. Using data mining algorithms for thalassemia risk prediction. *International Journal of Biomedical Science and Engineering*, 7(2):33–44, 2019.
[8] R. Galanello and A. Cao. Alpha-thalassemia. *Genetics in Medicine*, 13(2):83–88, 2011.
[9] Gretchen Holm and Kristeen Cherney. Available from https://www.healthline.com/health/thalassemia. pages 137–197, 7th August, 2018.
[10] Eyo J. E. Ivoke O. N. Nwani C. D. Odii E. C. Asogwa C. N. Ekeh F. N. Ivoke, N. and Atama. Anaemia prevalence and associated factors among women attending antenatal clinics in south-western ebonyi state, nigeria. international journal of medicine and medical sciences. *International Journal of Computer Theory and Engineering*, pages 1354–1359, 2013.
[11] Balkis A. Khair-allah, Mohammad Emad Rasmy, Mohammad A.M. Islam, Ahmad Darwish, and Mohammad M.S. Al-Haggar. Studying cor- relation between genotype and beta thalassemia major severity factors. *IEEE*, 2016.
[12] Y Lodhi. Economics of thalassemia management in pakistan. In S Ahmed, editor, *Thalassemia Awareness Week*. Friends of Thalassemia, 2003.
[13] T Lucke, S Pfister, and M Durken. Neurodevelopmental outcome and haematological course of a long-time survivor with homozygous alpha- thalassaemia: case report and review of the literature. *Acta Paediatr*, 94:1330–1333, 2005.
[14] Nishi Madan, Satendra Sharma, S. K. Sood, Roshan Colah, and H. M. Bhatia. Frequency of β-thalassemia trait and other hemoglobinopathies in northern and western india. *Indian Journal of Human Genetics*, 16(1):16–25, 2010.
[15] Jr. Muncie, H. L. and J. S. Campbell. Alpha and beta thalassemia. *American Family Physician*, 80(4):340–341, 2009.
[16] K. O. Osungbade and A. O. Oladunjoye. Preventive treatments of iron deficiency anaemia in pregnancy: a review of their effectiveness and implications for health system strengthening. journal of pregnancy 2012. pages 1–7, 2012.
[17] MU Rahman and Y Lodhi. Prospects and future of conservative management of beta thalassemia major in a developing country. *Pak J Med Sci*, 20:105–112, 2004.
[18] DP Steensma, V Viprakasit, A Hendrick, et al. Deletion of the alpha-globin gene cluster as a cause of acquired alpha-thalassemia in myelodysplastic syndrome. *Blood*, 103:1518–1520, 2004.
[19] World Health Organization (WHO). Who vitamin and mineral nutrition. geneva, switzerland: World health organization. 2011.
[20] World Health Organization (WHO). Nutritional anemia: report of a who scientific group. geneva,switzerland: World health organization. 2017.
[21] Nacholamp; Vichittumaros Kanjanaamp; Winichagoon Praneeamp; Fucharoen Suthat. Wongseree, Waranyuamp; Chaiyaratana. Thalassemia classification by neural networks and genetic programming. information sciences. 2017.
[22] M. M. Yassin et al. Genotype-phenotype characteristics of 13 tha- lassemia children in the gaza strip, palestine. *J Genet Disor Genet Rep*, 2:1–6, 2013.