Future Computing and Informatics Journal

Volume 5
Issue 1 (2020) *Volume 5 Issue 1*

Article 4

2020

Recent Advances and Machine Learning Techniques on Sickle Cell Disease

Noorh H. Alharbi
Computer Science Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia, nalharbi0373@gmail.com

Rana O. Bameer
Computer Science Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia, rbameer10@gmail.com

Shahad S. Geddan
Computer Science Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia, shahad.geddan@gmail.com

Hajar M. Alharbi
Computer Science Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia, hmsalharbi@kau.edu.sa

Digital Commons™ ELSEVIER

Future Computing and Informatics Journal

Homepage: https://digitalcommons.aaru.edu.jo/fcij/

Recent Advances and Machine Learning Techniques on Sickle Cell Disease

Noorh H. Alharbi[1,a], Rana O. Bameer[1,b], Shahad S. Geddan[1,c], Hajar M. Alharbi[1,d]

[1]Computer Science Department, Faculty of Computing and Information Technology,
King Abdulaziz University, Jeddah, Saudi Arabia

[a] nalharbi0373@gmail.com, [b] rbameer10@gmail.com,
[c] shahad.geddan@gmail.com, [d] hmsalharbi@kau.edu.sa

ABSTRACT

Sickle cell disease is a severe hereditary disease caused by an abnormality of the red blood cells. The current therapeutic decision-making process applied to sickle cell disease includes monitoring a patient's symptoms and complications and then adjusting the treatment accordingly. This process is time-consuming, which might result in serious consequences for patients' lives and could lead to irreversible disease complications. Artificial intelligence, specifically machine learning, is a powerful technique that has been used to support medical decisions. This paper aims to review the recently developed machine learning models designed to interpret medical data regarding sickle cell disease. To propose an intelligence model, the suggested framework has to be performed in the following sequence. First, the data is preprocessed by imputing missing values and balancing them. Then, suitable feature selection methods are applied, and different classifiers are trained and tested. Finally, the performing model with the highest predefined performance metric over all experiments conducted is nominated. Thus, the aim of developing such a model is to predict the severity of a patient's case, to determine the clinical complications of the disease, and to suggest the correct dosage of the treatment(s).

Published by Arab Journals Platform, 2020

1. Introduction

Sickle cell disease (SCD) is a severe hereditary disease that critically influences the patient's quality of life because of red blood cell abnormality (Khalaf et al., 2017). SCD was first discovered in 1910 (Wailoo, 2017), and today about 5% of the human population are healthy carriers of a gene for either sickle cell anemia or thalassemia. Every year, it is estimated that more than 300,000 newborns with severe forms of these diseases are born globally, with the majority of new cases being born in low and middle income countries (WHO, n.d.). According to Rogers (Rogers, 2017), SCD is predominant among citizens of the following regions/countries: Africa, the Caribbean, India, the Middle East, and the Mediterranean. Specifically, studies have shown that SCD is a quite common genetic disorder in Saudi Arabia. In some areas, the carrier status of SCD varies from 1.4% to 2%, and reaches up to 27% (El-Hazmi et al., 2011). The Saudi premarital screening program has estimated the incidence of SCD in the adult population at 0.26% for SCD and 4.2% for the sickle cell trait, with the highest incidence observed in the Eastern province (approximately 1.2% for SCD, and 17% for the sickle cell trait) (El-Hazmi et al., 2011). The current therapeutic decision-making process includes monitoring symptoms and complications of the disease, then adjusting the treatment accordingly (Creary and Strouse, 2014). This process can result in serious consequences affecting the patient's quality of life and could lead to late medical interventions and irreversible disease complications. The current evaluation process is also difficult and time consuming for medical staff.

Pre-symptomatic prediction of the severity of SCD for patients is a complex problem (Quinn, 2016). Using artificial intelligence (AI) techniques, more specifically machine learning (ML) models, to predict the severity of SCD for diagnosed patients may assist physicians and clinicians in the therapeutic decision-making process by using data collected from patients diagnosed with SCD.

This paper reviews the systems currently applied ML techniques to interpret medical data regarding SCD. Most of these systems predict the severity of SCD for diagnosed patients, while the remaining (currently available) systems are also used to suggest the proper dosage of the treatment drug (hydroxyurea) to prevent any disease complications. All proposed systems apply AI, specifically ML techniques, in the medical field and can be used to help physicians with both diagnosis and drug prescription processes.

The ability to predict the severity of SCD in diagnosed patients could advance medical research regarding the treatment process, alter the therapeutic decision-making conditions for physicians, and drastically improve the quality of life for patients with SCD (Creary and Strouse, 2014).

The remainder of the paper is organized in the following manner. Section 2 presents a clinical background on SCD. Section 3 presents a technical background about ML algorithms used in the medical field. In addition, Section

4 provides a summary of recent literature about predicting the severity of SCD and the estimation of drug dosage. Finally, Section 5 states the conclusion of this paper and suggests future challenges.

2. Clinical Background

2.1. SCD Description

Blood is a red viscous fluid consisting of red and white blood cells, plasma, and platelets. It flows into blood vessels and arteries, and is produced in the bone marrow. The heart muscle pumps the blood to all parts of the body, transferring nutrients and oxygen to all the cells, allowing the body to carry out its necessary functions (MOH, n.d.).

Red blood cells are a group of cells that number up to four or five million cells per cubic millimeter of blood fluid. The function of the red blood cells is to transport gases. These cells are distinguished by the pigment of hemoglobin which they contain, making the blood red. They bind the elements of iron, protein, and hemoglobin with oxygen and then transport these nutrients to the parts of the body. Red blood cells are prone to certain diseases, including malaria, hemophilia, thalassemia, anemia, and sickle cell anemia (MOH, n.d.).

Hemoglobin is an iron-rich, red compound that provides blood its red color. Hemoglobin permits red blood cells to transmit oxygen from the human lungs towards all parts of the human body (Rogers, 2017). SCD is a hereditary disease that modifies red blood cells. Red blood cells are usually circular, but in SCD patients, the cells take on the abnormal form of a crescent. The disease makes the sickle-shaped cells sticky and causes them to build-up in small blood vessels producing slow blood flow, which prevents oxygen from traveling to the rest of the body. Sickle red blood cells live between 10 and 20 days, whereas normal cells survive up to 120 days (MOH, n.d.). Sickle cell anemia is a long-term disease that has a serious impact on the expectancy and quality of life for patients. There are many complications resulting from sickle cell anemia, and most of these complications require emergency medical intervention. The most prominent of these serious complications are: (i) Acute chest syndrome, (ii) avascular necrosis, (iii) organ damage such as liver, kidney, and spleen, (iv) priapism, (v) pulmonary embolism (a condition of increased blood pressure within the arteries of the lungs), (vi) stroke, (vii) and tissue damage (CDC, 2020).

The sickle cell gene travels from one generation to another in a hereditary pattern known as the autosomal recessive genotype. Hence, the parents must pass on the defective gene in order for the child to become infected. The person needs two copies of the gene to acquire the disease.

1. Both mother and father are trait carriers.
2. Both mother and father are infected.
3. One is infected and the other is a carrier (MOH, n.d.).

2.2. SCD Types

Despite the multiple types of SCD disease, they are all similar in symptoms, yet they differ in severity. There are four types of sickle cell

Published by Arab Journals Platform, 2020

anemia, depending on the mutation that affects the person's genes.

1. Hemoglobin SC disease
2. Hemoglobin SS disease
3. Hemoglobin disease SB 0 beta-zero thalassemia
4. Hemoglobin disease SB +

In addition, there are other types of disease, although very rare, such as hemoglobin SD, hemoglobin SE, and hemoglobin SO (Rogers, 2017). The three main types of SCD can be explained as follows. The first and most prevalent is the case called Hb SS, which occurs when a patient inherits sickle cell genes from both parents. The second one is called Hb SC, and occurs when the patient inherits the sickle cell gene (S) and another gene that is generated from an irregular form of hemoglobin (C). Finally, if the patient inherits both beta thalassemia and the sickle cell gene, then the disorder is called s-beta thalassemia (Khalaf et al., 2017).

2.3. Suggested Treatments of SCD

Most medical centers use manual methods, such as observing blood sample test results and/or using the patient's history, to estimate both the severity of the disease and to give the correct dosage of treatment to the patient. This method depends on the experience of doctors (MOH, n.d.). So far, there is no definitive cure for the disease, but there are treatments that can alleviate the pain and help prevent the problems associated with SCD (MOH, n.d.). Such treatments include:

1. Blood transfusions
2. Medication to relieve pain such as morphine
3. Folic acid supplements to strengthen healthy blood cells
4. Vaccination and antibiotics to prevent infection
5. Marrow transplantation
6. Home care and healthy living (very important factors for these patients)
7. Hydroxyurea

Research has proven the beneficial effects of the drug, hydroxyurea. It could reduce the rate of the disease by 50% and indicates continuous long-term benefits, including the hindrance of organ damage. This medication is given orally, once every day, and comes in liquid or capsule form. The drug works by modifying the disease phenotype, which helps to increase hemoglobin production and keeps red blood cells round and breakable. This allows the cells easy access to body parts. This treatment is one of the most successful treatments (Luzzatto and Makani, 2019), and it has been clinically proven to reduce the above-mentioned complications (Creary and Strouse, 2014).

3. Machine Learning Techniques in Medical Fields

In recent years, the use of AI in medicine has been discussed extensively in the literature. AI can apply powerful techniques to 'learn' from the huge volumes of available health care knowledge. It has also sparked an energetic debate on whether AI systems could potentially substitute human physicians in the future (Jiang et al., 2017). Currently, as a result of the vast quantity of data and the continuous improvement of developed analytical methods, many attempts have been made to build systems to assist physicians in various medical tasks, such as early detection, diagnosis of diseases, and recommending and

prescribing treatment. (Jiang et al., 2017). Before ML systems can be used in the medical field, they must be trained using data from clinical activities. Depending on the type of the patient's problem, data could be derived from diagnoses, patient treatment plans, laboratory tests, or other general patient information, such as age or sex. The machine learning algorithms can be divided into three main categories:

- Supervised learning: All data points are labeled. This process works on predictive modeling by detecting relationships and patterns between input and output.
- Unsupervised learning: This process works with unlabeled data, and is well known for feature extraction and clustering.
- Semi-supervised learning: In this process, some data points are labeled, making it a hybrid of the two types above-mentioned.
- Reinforcement learning: This is called an agent, and it functions in an iterative way. The agent learns from its experiences until it discovers the full range of all possible states in the environment.

In medicine, ML applications most frequently use supervised learning classifiers because they provide more significant results than unsupervised ones. In the preprocessing step, unsupervised learning can be used to identify patterns in data or reduce dimensionality, hence making the consequence step more effective. Various ML algorithms have been used in medical applications such as artificial neural network (ANN), support vector machine (SVM), random forest (RF), and logistic regression (LR). Figure 1 shows the common supervised learning techniques used in medical fields, and it visibly demonstrates that SVM and ANN are the most common types. However, this fact remains true only when limited to the three main data types (image, genetic, and electrophysiological) (Jiang et al., 2017).

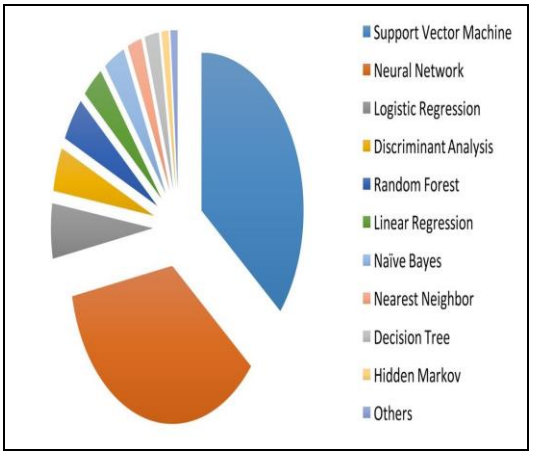

Figure 1. Most commonly used supervised learning techniques in medical fields (Jiang et al., 2017)

Next, the authors review the most widely-used supervised ML algorithms within medical applications. SVM is included because it is the most successful algorithm for accuracy in medical diagnosis (Goyal et al., 2018). In addition, RF is considered because it is the best performing algorithm in predicting SCD severity (Mohammed et al., 2019). Lastly, ANN algorithms are included because they have succeeded in various medical applications (Yasnitsky et al., 2015).

SVM Algorithm

SVM is one of the most notable supervised ML algorithms that can be employed to solve both regression and

Published by Arab Journals Platform, 2020

classification problems, but it is usually employed for binary classification. Finding a suitable hyperplane in an N-dimensional space (N — the number of features) that specifically classifies the data points is the objective of the SVM algorithm (Gandhi, 2018).

The main advantages of using SVM are its memory efficiency and its efficiency in high-dimensional spaces. On the other hand, this algorithm does not function well on large datasets or datasets with noise (Dhiraj, 2018). For SVM implementation, the most proven convex optimization techniques are readily available. The SVM algorithm has been widely applied in medical fields. It has been applied to achieve early detection of Alzheimer's disease and the diagnosis of cancer (Jiang et al., 2017).

RF Algorithm

Another commonly used supervised learning algorithm for both regression and classification is RF. The reason for naming the algorithm "random forest" is because the algorithm creates a "forest" from a number of decision trees (DTs), which are the basis for this algorithm (Chaudhary and Vasuja, 2019). DT is the portion that falls under supervised learning for classification algorithms. A tree-like structure is generated in this algorithm. The "tree" is traversed according to the conditions needed to find an outcome. Most researchers have used this technique because it is simple and lets them uncover small or large data structures and predict the value (Chaudhary and Vasuja, 2019).

ANN Algorithm

ANNs are a type of biologically inspired algorithms aiming to replicate the functionality of the human brain. They are considered powerful computational models that are capable of solving many complex tasks that involve generalization (Yasnitsky et al., 2015). These intelligent networks consist of a collection of connected artificial neurons, similar to biological neurons in the brain. These neurons are formed in layers, and different layers perform various transformations on their inputs.

- Input layer: The first layer in an ANN that receives raw input and represents the beginning of the network workflow.
- Hidden layer: This is used in multi-layered networks. It is the layer located between the input layer and the output layer, where neurons receive a set of weighted inputs and generate an output via an activation function.
- Output layer: The last layer in an ANN that generates the output of the network.

Inputs have to be multiplied by weights, which modifies the strength of the inputs, then the neurons will calculate the inputs to predict the final output. The weights in an ANN are modified in the training phase using learning algorithms.

Below are a few of the most popular types of ANNs:

- The feedforward neural network is the simplest implementation of an ANN. It is relatively easy to maintain due to its simplicity. It can also deal with noisy data, which is why ANNs are used in technologies like computer vision and face recognition (Mehta, 2019).
- The multi-layer perceptron (MLP) has three or more layers, and is a fully

connected network that is used for nonlinear problems. This type of ANN is heavily used in natural language processing and speech recognition (Mehta, 2019).

- The convolutional neural network applies multiple MLPs. It contains one or more convolutional layers that could be either pooled or completely interconnected (Jiang et al., 2017). It performs a convolutional operation on the input before passing the result to the next layer, and as a result, the network can be far deeper, yet with much fewer parameters (Mehta, 2019). Due to this ability, convolutional neural networks demonstrate exceptionally promising results in image classification, image and video recognition, and semantic parsing (Mehta, 2019).

4. Literature Review

SCD is phenotypically variable (Quinn, 2016). The ability to predict disease severity before patients develop any serious complications can have tremendous benefits for improving the general quality of life for patients and aid in the therapeutic decision-making process. This section reviews several related papers in order to compile comprehensive knowledge of the research. The selected literature was chosen based on the following criteria:

1. It focuses on SCD severity or suggests the appropriate dosage of hydroxyurea.
2. It was published from 2015 (inclusive) until present.
3. It employs ML techniques.

Most of the reviewed literature revolves around the prediction of the severity of the disease from various aspects, including organ damage, risk of anemia, and pain score. The remainder focuses on estimating the appropriate dosage of hydroxyurea for SCD patients with different machine learning approaches.

In both (Abd Al-Maaref et al., 2017) and (Khalaf et al., 2015a), the same intelligent monitoring system is proposed to monitor patient health relying on mobile apps installed on their smartphones. Whenever a patient enters his or her medical information, the developed expert system uses a ML classifier to determine whether the patient's situation is critical. The data used to train the classifier includes *patients carrying SCD traits* and *patients who are not carriers*. The difference between them is the classifiers used. In (Abd Al-Maaref et al., 2017), an additional sample was added to the dataset. This study (Abd Al-Maaref et al., 2017) tested six ML classifiers. These were JRip-Rules, Attribute Selected Classifier, Voted Perceptron, Bayes Net, Adaboost M1, and Logit boost (LB). LB achieved the best accuracy score (99.6%). In (Khalaf et al., 2015a), multiple experiments were carried out to identify normal patients from SCD patients using the following four ML algorithms: core vector regression, MLP, hyperpipes, and zero-rule based algorithms. The greatest accuracy (99.5984%) was achieved using the MLP classifier.

The authors (Ademola et al., 2018) used ensembles of ML models to predict the severity of SCD. The criteria used was *multiple physicians' opinions* in each case. The data was classified by experts into three classes: high risk, moderate risk, and low risk. The features also included laboratory tests. Seven ML models were tested including isolated classifiers, which are C4.5 DTs,

Published by Arab Journals Platform, 2020

Naïve Bayes, SVM, and ensembles of the isolated classifiers. The best performance was obtained by the ensemble method that combined the Naïve Bayes and C4.5 DTs classifiers.

In another study (Yang et al., 2018), the authors suggested a system that drew objective physiological measures from subjective self-reported pain scores by utilizing ML algorithms. The self-reported pain score for a patient, which ranges from 0 (no pain) to 10 (severe and unbearable pain), was included with each data point. The utilization of missing data was done using multiple imputation. The researchers recommended a pain prediction system for different situations: (i) inter-individual pain prediction with 11 pain scores, (ii) intra-individual pain prediction with 11 pain scores, and (iii) inter-individual pain prediction with condensed pain levels numbering less than 11 (6, 4 and 2 pain levels). From all these experiments and by using the performance measures, accuracy and F1 score, it was determined that the multinomial LR classifier provided the best performance measures compared to the remaining three classifiers used in this research (SVM, k-nearest neighbor and RF).

An early prediction of organ failure for patients with SCD was proposed by (Mohammed et al., 2019). Continuous physiologic data was collected on 63 adults based on 163 encounters. Several classification models were tested, including MLP, SVM, RF, and LR. Feature selection was used for extracting the features from each of the physiological data streams. The results showed that RF produced the highest accuracy (99.57%) in predicting organ failure prior to six hours before it began.

As has been discussed, several ML algorithms have been developed to assist healthcare professionals and physicians in estimating the severity of SCD using AI. This estimation is needed based on the severity of many serious complications that SCD causes. The authors reviewed research papers revolving around the prediction of the severity of the disease using different criteria. The most frequently tested algorithms were SVM and MLP, which indicates their popularity in the field. However, MLP and LB were the best-performing algorithm among all the reviewed papers. Table 1 summarizes all the research papers that were discussed in this section.

Despite the wealth of data on SCD, predicting the severity of the disease continues to be a complex issue, mainly because the disease is highly phenotypically diverse (Quinn, 2016). This attribute has made the problem of predicting the severity of SCD a good candidate for the use of AI, as is revealed in the above examples. Also, the authors are motivated to study this problem since few other attempts have been made to utilize AI to aid the therapeutic decision-making process.

Hydroxyurea is the effective drug used to treat SCD, as mentioned in Section 2.3. A number of studies (Khalaf et al., 2015b), (Khalaf et al., 2016), (Khalaf et al., 2017), have been proposed to help medical professionals and physicians predict the right dosage of hydroxyurea based on blood test results for SCD patients.

Table 1. Prediction of Severity in SCD Review

Author	ML Algorithms	Dataset	No of atients	No of Features	erformance Measure	est Performing ML Algorithm
halaf et al., 2015a)	– Core vector regression – Hyper pipes – MLP Zero-rule based	Two local spitals in the city of Liverpool	498	12	Accuracy	**MLP** 99.5984%
Abd Al-Maaref et al., 2017)	– Adaboost Ml – Bayes net – LB – Voted perceptron – JRip rules – Attribute lected classifier	Centre Caribéen de répanocytose	250	12	Accuracy	**LB** 99.5984%
(Ademola et al., 2018)	– C4.5 DT – NB – SVM	esley Guilds Obafemi Awolowo University Teaching Hospital Complex	NA	10	ccuracy & Precision	**Ensemble model (DT + NB)** with 86.957% Accuracy & 87.2% Precision
Yang et al., 2018)	– K-nearest neighbor Multinomial LR – RF – SVM	Duke University Hospital	40	6	ccuracy / F1 score	**Multinomial LR** 546 / 0.540 for 11-point scale rating .681 / 0.673 for 4-point scale rating .821 / 0.819 for 2-point scale rating
Mohammed et al., 2019)	– LR – MLP – RF – SVM	Iethodist Le Bonheur Hospital	63	5	Accuracy	**RF** 99.57%

A web-based system that aids healthcare physicians to provide SCD patients with a suitable amount of hydroxyurea was proposed using different types of ML algorithms (Khalaf et al., 2015b). In the initial stage of its proposed methodology, the system used a neural network consisting of a weighted model, where ANNs are accumulated into levels and the outputs from one network are catered into the subsequent in conjunction with desired outputs. In addition, it employed exploratory analysis using stochastic neighbor embedding plots. In the proposed system, several ML algorithms are used; however, the MLP algorithm outperformed the others by providing the lowest error rates.

An experimental study by (Khalaf et al., 2016) shows the use of different neural network models to predict the appropriate dosage of medication for SCD. The target dosage was divided into six bins. These models are voted perception (VP), back-propagation trained feed-forward neural network (BPXN), functional link neural network (FLNN), and radial-basis neural network (RBN) classifiers. The results for the models in order from best to worst using the area under the curve measure: BPXN: 0.989, FLNN: 0.972, RBN: 0.875, and the VP: 0.766 classifier. To demonstrate the importance of the previous models, researchers subsequently used a linear neural network as a baseline classifier, which produced an area of 0.849 and a random guessing model with an area of 0.524.

In addition, a group of researchers proposed a methodology that employed the holdout method to split the data (Khalaf et al., 2017). Then, the feature extraction method was conducted. Subsequently, the following models were used: SVM, RF, random oracle model, MLP trained using the Levenberg-Marquart learning algorithm (LEVNN), linear combiner network, Elman neural network, Jordan neural network, and the Elman-Jordan hybrid neural network. The target dosage of treatment was divided into three bins, and the highest accuracy/ F1 score was achieved using LEVNN.

Thus, this paper concludes a principle framework consisting of the following main steps:

1. Collecting, analyzing and preprocessing SCD patient data presented by a hospital or health care provider (including blood tests, disease severity levels, and suggested courses of treatment) to deal with missing values using various missing data imputation algorithms (Jakobsen et al., 2017; Yang et al., 2018).
2. Selecting the best features from the dataset by using a different filter, wrapper, or embedded method to improve system performance (Jain and Singh, 2018).

3. Applying different ML algorithms to the dataset to determine disease severity in newly diagnosed patients, clinical complications associated with the disease, and suggesting suitable patient treatment plan(s) based on disease severity.

5. Conclusion

Sickle cell anemia is a common yet severe hereditary disease that has a debilitating effect on patient life expectancy and quality due to complications caused by abnormal red blood cells. Since SCD was discovered, many medical advances have enabled disease treatment, increased survival rates, and improved patient quality of life. Early medical and therapeutic interventions can prevent many severe disease complications. Predicting the severity of SCD early in the diagnosis process can also make a remarkable difference in the diagnosed patient's quality of life, making treatment and drug prescription easier for medical professionals.

The authors of this paper performed a clinical review of SCD by describing the disease and discussing available treatment options with patients. The team also looked at similar work that attempted to estimate the severity of SCD in diagnosed patients to aid medical professionals in prescribing drugs. Such proposed systems help physicians, clinical pharmacists, and any other medical professionals who contribute to the diagnosis and drug prescription processes for SCD patients. This paper suggested a principle framework system that starts with a preprocessing step to estimate missing data. Dimensionality reduction is performed next to improve system performance. The results should then be tested with multiple widely-used ML algorithms to produce a state-of-the-art system that can prevent serious disease complications. System predictions of forthcoming SCD severity enables accurate estimating of hydroxyurea dosages, preventing complications. Future research should develop an advanced system incorporating an intelligent support system with ML algorithms, assisting physicians in making decisions that save time and enhance the quality of patient lives.

References

[1] Abd Al-Maaref, D., Alwan, J. K., Ibrahim, M., & Naeem, M. B., "The Utilisation of Machine Learning Approaches for Medical Data Classification and Personal Care System Management for Sickle Cell Disease," *2017 Annual Conference on New Trends in Information Communications Technology Applications (NTICT)*, Baghdad, Iraq, 2017.

[2] Ademola, B. J., Temilade, A., Chidozie, E. N., & Adebayo, I. P., "An

Published by Arab Journals Platform, 2020

Ensemble Model of Machine Learning Algorithms for the Severity of Sickle Cell Disease (SCD) among Pediatric Patients," *Computer Reviews Journal*, vol. 2, pp. 331-346, 2018.

[3] Alahmari, A. D., Aljurf, M., Alseraihy, A., Hamidieh, A. A., Alkindi, S., Rihani, R., Satti, T., Jastaniah, W., Alsaedi, H., Almohareb, F., Al-Jefri, A., & Rasheed, W., "Hematopoietic Stem Cell Transplantation for Patients with Sickle Cell Disease in the Eastern Mediterranean," *Hematology/Oncology and Stem Cell Therapy*, vol. 13, no. 2, pp. 106–110, 2020.

[4] CDC, "Complications and Treatments of Sickle Cell Disease," June 3, 2020. [Online]. Available: https://www.cdc.gov/ncbddd/sicklecell/treatments.html

[5] Chaudhary, D. & Vasuja, E. R., "A Review on Various Algorithms used in Machine Learning," *International Journal of Scientific Research in Computer Science, Engineering, and Information Technology*, vol. 5, no. 2, pp. 915–920, 2019.

[6] Creary, S. E. and Strouse, J. J., "Hydroxyurea and Transfusion Therapy for the Treatment of Sickle Cell Disease: A Pocket Guide for the Clinician," Washington, DC: American Society of Hematology, 2014.

[7] Dhiraj, K., "Top 4 Advantages and Disadvantages of Support Vector Machine or SVM," *Medium*, June 14, 2018. [Online]. Available: https://medium.com/@dhiraj8899/top-4-advantagesand-disadvantages-of-support-vector-machine-or-svm-a3c06a2b107

[8] El-Hazmi, M. A., Al-Hazmi, A. M., & Warsy, A. S., "Sickle Cell Disease in Middle East Arab Countries," *The Indian Journal of Medical Research*, vol. 134, no. 5, pp. 597–610, 2011.

[9] Fumo, R., "Types of Machine Learning Algorithms You Should Know," *Medium*, June 15, 2017. [Online]. Available: https://towardsdatascience.com/types-of-machine-learning-algorithmsyou-should-know-953a08248861

[10] Gandhi, R., "Support Vector Machine — Introduction to Machine Learning Algorithms," *Medium*, June 7, 2018. [Online]. Available: https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47

[11] Goyal, H., Khandelwal, D., Aggarwal, A., & Bhardwaj, P., "Medical Diagnosis Using Machine Learning," *BODH BPIT International Journal of Technology & Management*, vol. 4, no. I, pp. 7-11, 2018.

[12] IMGBIN, "Sickle Cell Disease Anemia Sickle Cell Trait Red Blood Cell," May 28, 2018. [Online]. Available: https://imgbin.com/png/2bfW7eq6/sickle-cell-disease-anemia-sickle-cell-trait-redblood-cell-png

[13] Jain, D. and Singh, V., "Feature selection and classification systems for chronic disease prediction: A review", *Egyptian Informatics Journal*, vol. 19, no. 3, pp. 179-189, 2018.

[14] Jakobsen, J., Gluud, C., Wetterslev, J., and Winkel, P., "When and how should multiple imputation be used for handling missing data in randomised clinical trials—a practical guide with flowcharts", *BMC Medical Research Methodology*, vol. 17, no. 1, 2017.

[15] Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., ... & Wang, Y., "Artificial Intelligence in Healthcare: Past, Present and Future," *Stroke and Vascular Neurology*, vol. 2, no. 4, pp. 230-243, 2017.

[16] Khalaf, M., Hussain, A. J., Al-Jumeily, D., Keenan, R., Fergus, P., & Idowu, I. O., "Robust Approach for Medical Data Classification and Deploying Self-care Management System for Sickle Cell Disease," *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, Liverpool, UK, pp. 575-580, 2015a.

[17] Khalaf, M., Hussain, A. J., Al-Jumeily, D., Keenan, R., Keight, R., Fergus, P., & Idowu, I. O., "Applied Difference Techniques of Machine Learning Algorithm and Web-based Management System for Sickle Cell Disease," *2015 International Conference on Developments of E-Systems Engineering (DeSE)*, Dubai, United Arab Emirates, 2015b.

[18] Khalaf, M., Hussain, A. J., Al-Jumeily, D., Keight, R., Keenan, R., Fergus, P., ... Idowu, I. O., "Training Neural Networks as Experimental Models: Classifying Biomedical Datasets for Sickle Cell Disease," *International Conference on Intelligent Computing*, Cham: Springer International Publishing, pp. 784-795, 2016.

[19] Khalaf, M., Hussain, A. J., Keight, R., Al-Jumeily, D., Fergus, P., Keenan, R., & Tso, P., "Machine Learning Approaches to the Application of Disease Modifying Therapy for Sickle Cell Using Classification Models," *Neurocomputing*, vol. 228, pp. 154-164, 2017.

[20] Luzzatto, L. and Makani, J., "Hydroxyurea — An Essential Medicine for Sickle Cell Disease in Africa," *New England Journal of Medicine*, vol. 380, no. 2, pp. 187-189, 2019.

[21] Mehta, A., "A Comprehensive Guide to Types of Neural Networks," *Digital Vidya*, January 25, 2019. [Online]. Available: https://www.digitalvidya.com/blog/types-of-neural-networks

[22] MOH, "Hematology - Sickle Cell Anemia," no date. [Online]. Available: https://www.moh.gov.sa/HealthAwareness/EducationalContent/Diseases/Hematology/Pages/SickleCell-Anemia.aspx

[23] Mohammed, A., Podila, P. S., Davis, R. L., Ataga, K. I., Hankins, J. S., & Kamaleswaran, R., "Machine Learning Predicts Early-onset Acute Organ Failure in Critically Ill Patients with Sickle Cell Disease," *bioRxiv*, pp. 614941, 2019.

[24] Quinn, C. T., "Minireview: Clinical Severity in Sickle Cell Disease: The Challenges of Definition and Prognostication," *Experimental Biology and Medicine*, vol. 241, no. 7, pp. 679-688, 2016.

[25] Rogers, G., "Sickle Cell Anemia," *Healthline*, March 29, 2017. [Online]. Available: https://www.healthline.com/health/sickle-cell-anemia

[26] Salem, R. and Abdo, A., "Fixing rules for data cleaning based on conditional functional dependency," *Future Computing and Informatics Journal*, vol. 1, no. 1, pp. 10-26, 2016.

[27] Wailoo, K., "Sickle Cell Disease—A History of Progress and Peril," *New England Journal of Medicine*, vol. 376, no. 9, pp. 805-807, 2017.

[28] WHO, "Human Genomics in Global Health," no date. [Online]. Available: https://www.who.int/genomics/public/geneticdiseases/en/index2.html

[29] Yang, F., Banerjee, T., Narine, K., & Shah, N., "Improving Pain Management in Patients with Sickle Cell Disease from Physiological Measures Using Machine Learning Techniques," *Smart Health*, vol. 7-8, pp. 48–59, 2018.

[30] Yasnitsky, L. N., Dumler, A. A., Poleshchuk, A. N., Bogdanov, C. V., & Cherepanov, F. M., "Artificial Neural Networks for Obtaining New Medical Knowledge: Diagnostics and Prediction of Cardiovascular Disease Progression," *Biology and Medicine (Aligarh)*, vol. 7, no. 2, BM, 2015.

[31] Yiu, T., "Understanding Random Forest: How the Algorithm Works and Why it is so Effective," *Medium*, June 12, 2019. [Online]. Available: https://towardsdatascience.com/understanding-random-forest-58381e0602d2