Mathematical Biosciences and Engineering

MBE, 19(10): 10060–10077.
DOI: 10.3934/mbe.2022470
Received: 11 April 2022
Revised: 28 June 2022
Accepted: 30 June 2022
Published: 15 July 2022

http://www.aimspress.com/journal/MBE

Research article

A mathematical modelling to detect sickle cell anemia using Quantum graph theory and Aquila optimization classifier

P. Balamanikandan and S. Jeya Bharathi

Department of Mathematics, Thiagarajar College of Engineering, Madurai, Tamilnadu, India

* **Correspondence:** Email: mailtobalamaths@gmail.com, sjbmat@tce.edu.

Abstract: Recently genetic disorders are the most common reason for human fatality. Sickle Cell anemia is a monogenic disorder caused by A-to-T point mutations in the β-globin gene which produces abnormal hemoglobin S (Hgb S) that polymerizes at the state of deoxygenation thus resulting in the physical deformation or erythrocytes sickling. This shortens the expectancy of human life. Thus, the early diagnosis and identification of sickle cell will aid the people in recognizing signs and to take treatments. The manual identification is a time consuming one and might outcome in the misclassification of count as there is millions of red blood cells in one spell. So as to overcome this, data mining approaches like Quantum graph theory model and classifier is effective in detecting sickle cell anemia with high precision rate. The proposed work aims at presenting a mathematical modeling using Quantum graph theory to extract elasticity properties and to distinguish them as normal cells and sickle cell anemia (SCA) in red blood cells. Initially, input DNA sequence is taken and the elasticity property features are extracted by using Quantum graph theory model at which the formation of spanning tree is made followed by graph construction and Hemoglobin quantization. After which, the extracted properties are optimized using Aquila optimization and classified using cascaded Long Short-Term memory (LSTM) to attain the classified outcome of sickle cell and normal cells. Finally, the performance assessment is made and the outcomes attained in terms of accuracy, precision, sensitivity, specificity, and AUC are compared with existing classifier to validate the proposed system effectiveness.

Keywords: monogenic disorder; genetic disorder; sickle cell anemia; red blood cells; hemoglobin; Aquila optimization based cascaded LSTM classifier; elasticity property features; Quantum graph theory model

1. Introduction

Monogenic disorders are usually caused by the single gene lesion predominantly, though the manifestation of phenotype might depend on several extents on the extra genetic variants in same or other genes, changes in epigenetic and environmental conditions. Though most of the individual monogenic diseases are rate, altogether they signify a high disease load among the population [1]. The DNA arrays dependent tests are presently available for diagnosing the human genetic disease which includes tests that were designed for searching the disease that cause mutations in any one of gene numbers that were known to be mutated at desired disease kinds for example the disease that are associated with ataxia [2,3]. Also, a technique of sequencing generation might facilitate the entire exome analysis or the genome at some individual. By data collection from huge group of healthy individuals of population followed by data archiving in publications, bioinformatic resources which were established for the assessment of likelihood that a specific genomic variant or sequence variant present in specified patient is of pathologic importance.

Typically, Sickle Cell Disease (SCD) is regarded as a hereditary form of anaemia which outcome from single mutation at 6^{th} codon of β-globin chain (from the glutamic acid to the values) of adult tetramer ($\alpha_2\beta_2$), haemoglobin (Hb) which is more prone to the polymerization of lower levels of oxygen. This is one such sever and prevalent monogenic disorder and several million around the world and over 100,000 individuals in United States were much affected by these both chronic and acute SCD manifestations like crises of frequent pain, infarct of silent cerebral, early death, and end organ damage [4]. Also, previous studies shows that about 7000–10,000 children having Major Thalassemia are born in India every year. The treatments that reduce the complications and symptoms of SCD like transfusions of blood, preventive therapies that include penicillin prophylaxis and pneumococcal vaccination, and therapy of hydroxyurea were leverages in clinics. However, blood transfusion might not correct the phenotype and outcome in overload of iron once not accompanied by the therapy of aggressive chelation. The treatment of Hydroxyurea offers clinical significance over the fetal globin induction ($\alpha_2\gamma_2$, HbF) that competes sickle globin, thereby decreasing the symptoms of SCD, however the response to the hydroxyurea is not a uniform one among the patients that concerns for the long-term use remain contempt abundant indication for safety [5].

For the monogenic disorders, various non-invasive prenatal diagnosis (NIPD) approaches were presented. As the cell free fetal DNA cffDNA occurs in a maternal cfDNA background, previous researchers, like the autosomal diagnosis disorders, genetic disorders that are inherited paternally like rhesus and a compound heterozygotes exclusion of the autosomal recessive disorders such as Maternal plasma, β-thalassemia comprise of cffDNA mixture and the maternal cfDNA though fetal alleles half were maternally inherited. Consequently, a quantitative technique for counting the precise number of various alleles is thus needed for deducing the given locus genotype. For this reason, various techniques were presented in the series study that includes relative mutation dosage (RMD) and a relative haplotype dosage (RHDO) analysis. The approach of RMD thus counts the number of DNA molecules directly for sorting the mutant ratio and wild-type alleles with the use of digital polymerase chain reaction (PCR). The RHDO approach principle is to infer the inheritance of maternally transmitted fetal mutations on quantifying the relative haplotypes dosages along with a single nucleotide polymorphism (SNP) that allies in and around the specified or targeted genes. The data regarding the parental haplotypes must initially be attained by the sequencing techniques whereas the prior knowledge of proband was anymore not needed. Typically, on the whole both RHDO and RMD

techniques were found to be helpful at-risk couples of monogenic disorders who might request for NIPD for their first child [6]. The existing techniques for diagnosing the conditions of monogenic disorder may include short-read sequencing of exomes or the genomes [7].

Though the yield of diagnosis from these techniques is a promising one that ranges from 26% to 40% which might leave cases that are unresolved. The yield is thus augmented by the reanalysis in contradiction of newly discovered disease allied genes and variants, or on employing the analysis that are family based for identifying the variants of de novo, however the enhancements are modest [8,9].

The motivation behind this work is that, the direct recognition or the read out of the DNA bases by means of DNA-binding protein includes the amino acids which directly interacts with the specific features to each base. The experimental validation also indicates that in several cases the protein attains the partial sequence through indirect recognition that is by recognizing the structural properties of DNA. Though there were several optimization techniques employed along with classifier like GA, Short Term Load Forecasting (STLF), with fuzzy logic approach. But the computational time is high in these traditional approaches [10–12]. Also, there were several optimization techniques like arithmetic optimization approach [13], Dwarf Mongoose optimization [14], Reptile Search Algorithm (RSA) [15], and Ebola Optimization Search Algorithm [16]. In this paper, Aquila Optimization [17] is presented. The main intention of the proposed work is to present a mathematical model using Quantum graph theory to extract elastic properties and to distinguish them as normal cells and sickle cell anemia (SCA) in red blood cells.

The residual part of the paper is arranged as given below. Section II provides the related works. The proposed work is described in Section III. The analysis of performance and evaluation with the other techniques is given in Section IV. Finally, Section V offers the conclusions.

2. Related works

Reference [18] projected on the statistical analysis of genetic data to identify the neuro developmental disorder and also monogenic disorders. One of the neurodevelopmental disorders is because of sporadic de novo variants that occurs rarely but affects severely. As the genomic screens of large-scale are absent, epidemiological disease recognition is impossible for the variety of de novo monogenic variants. It is important for the researchers and clinicians to know about the incident of disease occurrence for planning the treatment process. So, the estimates of calculated validity are proceeded with two corroboration analyses.

Reference [19] proposed about the technique of analyzing DNA of parents and also the complex mutation assessments were illustrated. By this technique the multiplexing capabilities of finding mutations and number of genes would be calculated simultaneously. The detection of fetal monogenic diseases is agreeable and can be done by analysis of maternal plasma DNA. The monogenic disease analysis in prenatal fetus is highly scalable by the scheme.

Reference [20] focused on the field of monogenic disorders that is interrelated with NIHF (nonimmune hydrops fetalis). The gathering of fluids in the fetal compartments is dangerous to the life of fetus and this is said to be hydrops fetalis. Many molecular tests are performed in pregnancies of hydrops fetalis and analyzed. Many monogenic disorders have impact because of nonimmune hydrops fetalis.

Reference [21] demonstrated about neurodevelopment disorders are the effects of genetic variations that risks more in society. Several human disorders occur by solitary protein coding genetic variant. In

general, it is predicted that phenotypic characters of a person having monogenic disorder is impacted by the similar variants, in associating with intracranial volume, autism and birth weight and height.

Reference [22] estimated the monogenic variants role over the dystonic disorders range on offering guidance for the exploitation of personalized strategies care and fostering follow-ups for pathophysiological explorations. From the blood genomic DNA is extracted and sequenced the entire exome. Among the categories, 104 out of 160 were the noticed variants influenced genes that are related to neuro development disorders.

Reference [23] performed whole exome sequencing in a multi-center cohort consisting of 114 families involving 138 affected with CKD, in order to regulate the similarity of detection of monogenic reasons of CKD in adults. Nearly 500 monogenic disorders of chronic kidney disease CKD have been found in pediatrics. Affected people were employed from 16 families containing extra renal characteristics, positive family past history of 78 families and without family past history and extra renal characteristics of 20 families. Here it is found that a pathogenic mutation is in a CKD gene among 42 families out of 114. Among the identified 42 families with monogenic, the clinical diagnosis in 17 is confirmed by WES, recognized a diagnosis in 16 families with CKD and corrected the 9 clinical diagnoses.

Reference [24] analysed sequencing exome and 36 children and adults' data variants were analysed using Common variable immunodeficiency with strong relationships to evaluate the monogenic amount. Recently CVID of 20 percent cases have a defined genetic reason. Diagnosis is on the basis of immunological norms and clinical, after eliminating the other diseases that cause same phenotypes.

Reference [25] studied that monogenic lupus is a kind of systemic lupus erythematosus SLE that presents in single gene defect patients. An important gene numbers have been considered in monogenic lupus offering insights into a process of complex disease. Here the genes related to monogenic lupus will be explained and the work of pathogenic influenced by the mutations included by stimulating the autoimmunity. Although the monogenic lupus considers only a small group of lupus patients, it offers important insight into the lupus mechanism and approaches of treatment.

Reference [26] focused to recognize clinically related variants within the 68 monogenic inflammatory bowel disease IBD genes in an unchosen cohort of IBD. Monogenic inflammatory bowel disease (IBD) consists of rare cause of mendelian gut infection usually occurs in infants with severe disease. Here, whole exome sequencing was conducted on pediatric onset disease patients.

Reference [27] recognized formerly described homozygous mutation in Complement C1q C Chain gene, that describes the patient with SLE phenotype and immunodeficiency. They also found heterozygous variants in *DNASE1L3, C1S, DNASE1, RNASEH2A* and *IFIH1* and 7 rare genes that are included in monogenic Systemic lupus erythematosus. Their experimental findings denote a complex involvement to the entire gene risk of SLE by the occasional gene variants related with the SLE monogenic forms.

Reference [28] attempted to find the monogenic disorders which distribute the characteristics of Rett syndrome RTT. RTT is an early occurring neuro developmental disorder which is caused by MECP2 gene mutations. But other gene CDKL5 and FOXG1 defects can cause performances that are similar to classic RTT though they are not entirely similar. 437 patients were researched with RTT like and in 24,217 genes associated to RTT phenotype was performed through a system of HaloPlexTarget-Enrichment.

Reference [29] revealed three mutations in COL1A2, LRP5 and COL1A1 in genetic analysis. The results in the data showed that formerly unidentified monogenic bone disorders play a significant role in Pregnancy-associated osteoporosis. Pregnancy is regarded as a skeletal risk factor to promote the starting onset of skeletal monogenic disorders thereby describing the cases with osteoporosis related to pregnancy.

Reference [30] reviewed the abnormalities of methylation that is relevant to monogenic diseases, their association with phenotype and genotype and the usage limitation and output interpretation. The modification of Normal methylation is done at multiple or single loci as a result of gene variants acting in trans or cis or through communication with environmental factors. Particular patterns of methylation or signs can be found in DNA for several developmental disorders.

Though, there were several techniques employed for the detection of sickle cell anemia, there were some limitations like low detection accuracy, computational complexity, and fails to analyze the elasticity properties. To overcome this, the proposed model is suggested.

3. Proposed work

A detailed elucidation of the proposed work is provided in this section. The entire flow of the proposed work is given below in Figure 1. At first, input DNA sequence is considered and the features of elastic property are thus extracted with the use of Quantum graph theory model at which the spanning tree formation is made followed by construction of graph and Hemoglobin quantization. Then, the extracted properties are optimized using Aquila optimization and classified using cascaded LSTM to attain the classified outcome of normal cells and sickle cell.

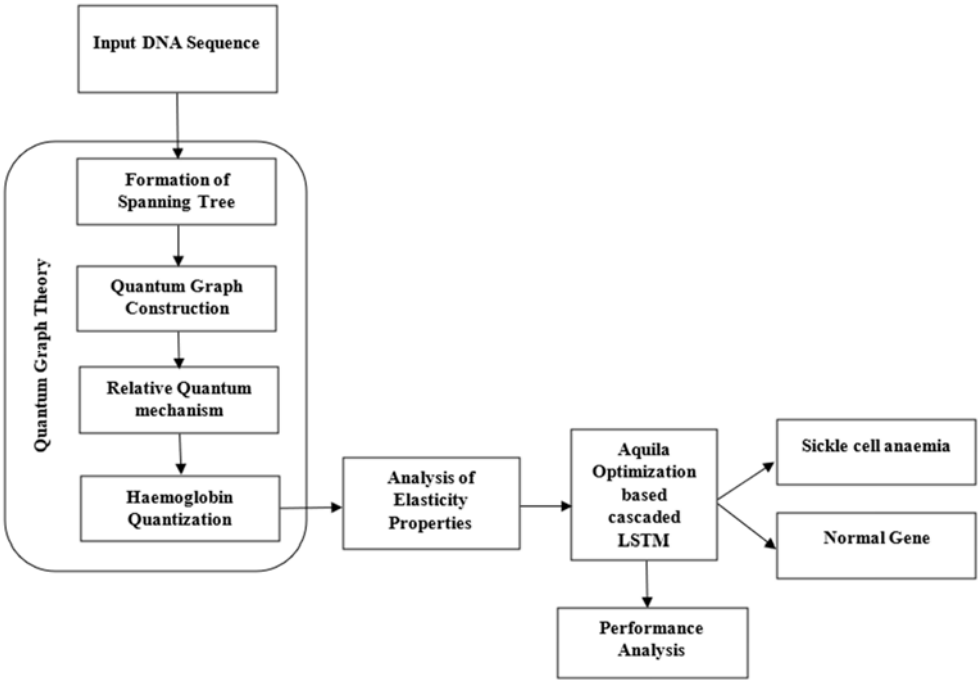


Figure 1. Overall Flow depiction of proposed model.

3.1. Input DNA sequence

The input DNA sequence from the input dataset is taken and the elastic property features are extracted by using Quantum graph theory model. After which, the extracted properties are optimized and classified to attain the classified outcome of sickle cell and normal cells in the input DNA sequence.

3.2. Quantum graph theory to extract elastic property features

The input DNA sequence is considered as input and Quantum graph theory is employed for the extraction of elastic property features at which the spanning tree is formed by the construction of Quantum graph. The relative quantum mechanism is computed followed by the quantization of Hemoglobin. To form a metric, graph a quantum one, a differential Hamiltonian operator on δ, which is sometimes required to be self-adjoint.

The negative second derivative,

$$f(y) \to -\frac{d^2 f}{dy^2} \tag{1}$$

The more general Schrodinger operator,

$$f(y) \to -\frac{d^2 f}{dy^2} + V(y)\, f(y) \tag{2}$$

It can rewrite as follows,

$$f(y) \to \left(\frac{1}{i}\,\frac{d}{dy} - A(y)\right)^2 + V(y)\, f(y) \tag{3}$$

At a later stage, the possibility of more general scalar or matrix differential or pseudo-differential operators will be discussed. We will focus on the most frequent scalar second order differential operators here, and for the sake of clarity in Eq (1), the Eq (2) is formed.

Vertex circumstances: According to the standard Sobolev trace theorems, a function $f \in Hn^2(e)$ and its first derivative have correctly defined values at the edge e's endpoints. The vertices of the second order derivatives of Hn^2 functions are already devoid of traces. As a result, only the values of f and $\frac{df}{dy}$ at a vertex v may be included in the vertex criteria.

The criteria can, in theory, combine the values at various vertices. It will, however, focus on local vertex conditions exclusively for the time being, that is, those that involve the values of functions and their derivatives at a single vertex at a time.

We'll stick to finite graphs here to avoid certain technical details. We assume, in other words, that the number of edges |E| (and thus the number of vertices |V|) is finite. The length of the edges can still be unlimited.

A typical Kirchhoff's vertex condition,

$$\begin{cases} f(y)\ is\ continuous\ on\ \delta \text{ and} \\ at\ each\ vertex\ v\ one\ has\ \sum_{e\,\in EE_v} \frac{df}{dy_e}\ (v) = 0 \end{cases} \tag{4}$$

Sum of overall edges e incident to the vertex v and the derivatives are taken in the orders gone from vertex.

It's also natural since the condition of Neumann boundary for the Laplacian operation. As we'll see in future, the quadratic form domain of the corresponding operator does not enforce any restrictions on a role other than that it be in $Hn^1(\delta)$ (and thus continuous). Another point worth mentioning is

that under the boundary requirements, all vertices of degree 2 can be eliminated, resulting in two smooth edges.

This is adequate for handling the challenge of describing the criteria for a junction of single d at a vertex v edges (a "star graph") given local vertex conditions. Because our operator functions as a second order operator along each edge, one would expect to begin two circumstances for each edge, and so the number of conditions at each vertex must correspond to the vertex's degree d.

As previously stated, the conditions for functions in H2 on each edge may only include the function's border values and derivative. The most generic (homogeneous) form of such a condition is plainly,

$$X_v\, F(v) + Y_v F'(v) = 0 \tag{5}$$

Here, X_v and Y_v are $r * r$ matrices, $F(v)$ is a column vector of the values at the mentioned vertex v, that function is follows,

$$F'(v) = \big(f_1(v), \dots, f_r(v)\big)^t \tag{6}$$

The derivatives considered along with these edges at the outgoing direction. The rank of the $r \times 2r$ matrix (X_v, Y_v) must be equal to r so as to safeguard the conditions that are independent.

$$\text{T}=f_1(v) \tag{7}$$

$$\text{U}=f_r(v) \tag{8}$$

Given two vertexes T and U, sequence UT is considered, with their history. By description, a components number essential for building T once appended to U is represented as $o(UT) - o(U)$. This is less than or equal to $o(T)$ since at any specified step of the making procedure T (on building the UT sequence) for this a huge search space is exploited because of the U existence. Hence, the process of copying is longer which could decrease the exhaustive components number. It could be observed from sub-additivity of complexity of Lempel-Ziv $o(UT) \leq o(U) + o(T)$. In what way $o(UT) - o(U)$ will be less than o(T) might depends on similarity degree among U and T. For Instance, assume $U = GGA\ CTT\ CTT, R = GCA\ GTT\ CTT\ and\ T = GGA\ TTT\ CCT$. The comprehensive antiquities of these arrangements might be signified as:

$$h_E(U) = \text{G. GA. CT. TC. TT} \tag{9}$$

$$h_E(R) = \text{G. CA. G. TT. CTT} \tag{10}$$

$$h_E(T) = \text{G. GA. TT. TC. CT} \tag{11}$$

On yielding $o(U) = o(R) = o(T) = 5$. The comprehensive sequences histories UT, and RT might be:

$$h_E(\text{UT}) = \text{G. GA. CT. TC. TT. GGA. TTT. CCT} \tag{12}$$

$$h_E(\text{RT}) = \text{G. CA. G. TT. CTT. GGA. TTTC. CT} \tag{13}$$

Correspondingly, we observe that it might take about three steps for building T at the process of UT production. Instead, five steps are employed for generating T at the process of production of the RT. The reason is that it could take some steps in next case as T is nearer to 'S' on comparing 'R'. for

instance, it was observed that on looking the CTT and GGA patterns which S and T shares. The number of steps is formulated as it takes some steps for generating a T sequence S by $c(UT) - c(U)$. Therefore, if S is nearer to T on comparing R, then it would be expected that $c(UT) - c(U)$ to be lesser than $c(RT) - c(R)$ since it is the case in above instances. Depending on the closeness idea, four distance measures were computed.

Distance metrics 1: Two sequences U and T are given, describe the $d\ (U,T)$ function as:

$$dist(U,T) = max\{o(UT) - o(U), o(TU) - o(T)\} \tag{14}$$

So as to eradicate the length effect on distance measure, a more substantial role might be the standardized form of $dist(\cdot,\cdot)$:

Distance metrics 2: Two sequences S and Q are given, and the function is defined as:

$$dist^*(U,T)\ as\ dist^*U,T) = \max\{\mathrm{o(UT)} - \mathrm{o(U)}, \mathrm{o(TU)} - \mathrm{o(T)}\} \max\{\mathrm{o(U)}, \mathrm{o(T)}\} \tag{15}$$

One more distance measure which might obviously follow from the building sequence T idea using U is signified as the 'sum distance'. This term is employed at this sense which accounts for total amount of steps it would take build T from U and vice versa.

Distance metrics 3: Two sequences U and T are given that defines the $d1(U,T)$ function as:

$$d1(U,T) = o(UT) - o(U) + o(TU) - o(T) \tag{16}$$

Likewise, $dist1(\cdot,\cdot)$ normalized version could be expressed as shown:

Distance metrics 4: Two sequences S and Q are given, which describe the function of $d * 1\ (S,Q)$ as$*$

$$dist^{1*}1\ (T,U) = o(TU) - o(T) + o(UT) - o(U) \tag{17}$$

A distance metric, $dist(\cdot,\cdot)$, might gratify the subsequent circumstances:

- $D(U,T) \geq 0$ at which the equality is satisfied if $U = T(identity)$.
- $D(U,T) = D(T,U)\ (symmetry)$.
- $D(U,T) \leq D(U,T) + D(R,T)$ (Triangle inequality). Appropriate for a metric to be an effective evolutionary change measure it might also satisfy these conditions:
- $D(T,R) + D(U,R) \leq max\{D(T,U) + D(R,T), D(T,R) + D(U,R)$ (additivity).

Consequently, effective distance metrics fourth condition are tested on comparing the matrix of distance created by presented metrics with the reconstructed branch lengths of resultant tree.

Theorem (Elastic co-efficient estimator): Let δ be a metric graph with finite edges, Consider the operator Hn working as $-\frac{d^2}{dy^2}$ on each edge e, with the domain including functions belonging to Hn^2 and specific local vertex conditions involving function vertex values and derivatives. If and only if the vertex conditions can be stated in one (and hence any) of the three forms below, the operator is self-adjoint.

- Equation (14) at each vertex, where $\{X_v, Y_v \mid v \in V\}$ for the collection of matrices of size $r_v\ x\ r_v$ such that,

The maximal rank of (X_v, Y_v) with $r_v\ x\ 2r_v$ matrix size.

The matrix (X_v, Y_v^*) is self-adjoint.

- There exists a unitary $r_v \; x \; r_v$ matrix U_v for every vertex v of degree r_v such that the vertex criteria at v are,

$$i(U_v - \beta)F(v) + (U_v - \beta)F'(v) = 0 \tag{18}$$

where β is the $r_v \; x \; r_v$ identity matrix.

The α_2 value associated with protein-induced DNA alterations from Y_v to the X_v method. The elastic coefficients estimated have the following assets. The tilt modulus ratio F_{11}^{XY} to the roll modulus F_{22}^{XY} is at range, and twist modulus ratio F_{33}^{XY} for the harmonic mean,

$$A^{XY} = 2\left[\frac{1}{F_{11}^{XY}} + \frac{1}{F_{22}^{XY}}\right]^{-1} \tag{19}$$

The formation of spanning tree and the outcomes of graph constructed by spanning tree model for each attribute is represented below.

The attributes like HB, HCT, MCV, MCH, MCHC, RDW, and RBC count are considered for the construction of graph using quantum graph theory model. The graph construction takes place based on the variable range values of each attribute and is shown in Figure 2. The attributes values should be within the normal range, if there occurs some variation, then there will be some disorder. Hemoglobin is a protein that carried oxygen in red blood cells. If the hemoglobin becomes defective it causes the red blood cells to change its shape causing sickle cell anemia. The healthy range of hemoglobin (for men are 13.2–16.6 grams per deciliter, for women it is 11.6–15 grams per deciliter (dL)). The normal range for HCT is (38.3–48.6 percent for men and 35.5–44.9 percent for women). The normal MCV value should be between 80 and 100 fL. In case of MCH, the normal value is 29 ± 2 picogram per cell. MCHC denotes the amount of hemoglobin per unit volume and the normal range lies between 33.4 and 35.5 grams per dL. If low MCHC is found then it is considered as anemia due to deficiency of iron and it indicates the thalassemia. The healthy range of RDW is 12.2–16.1 percent in female and it is 11.8–14.5 percent in males. If the score value out of this range occurs, then there is a nutrient deficiency, infections or the other disorder. Usually, women have lower RBC count than men, and the RBC level tends to decline with respect to age. The normal range lies between 4.7 and 6.1 million cells per microliter (cells/mcL) for men and for women it is 4.2–5.4 million cells/mcL.

3.3. Aquila optimization based cascaded LSTM

Once the quantization is attained by means of Quantum graph theory, the elasticity properties are analyzed and are optimized so as to enhance the performance of classifier [17,2]. The Aquila optimization based cascaded LSTM approach is employed so as to classify the monogenic disorder sickle cell anemia and normal cell. The algorithm for this Aquila Optimization based cascaded LSTM is shown below:

Algorithm 1: Aquila optimized cascaded LSTM classification
Input: Train features $Train$, Test Feature $Test$, train Label $Label$
Output: Predicted Class $Output_{cl}$
Step 1: The encoder has four convolutional layers, each of which estimates feature maps with different kernel sizes (3, 3, 5, 5).

$$I_{conv} = \left[\frac{Train+(2*p_c)+k_c}{s_c}\right]$$

I_{conv}—Output convolutional features
p_c—Convolution padding size
k_c—Convolution kernel size
s_c—Convolution stride size

Step 2: Three max pooling stages are applied; the signal sizes are divided into two equal parts based on the max pooling process.

$F_{max} = \max(I_{conv}(j), G_F)$

where $G_F = \frac{\delta l}{\delta k_i} * I_{conv}(i)$

δl is the length of the estimated convolutional features
δk_i is the derivative parameter computed by using multi stage probability sampling

$$\delta k_i = \frac{N_I \sigma_I^2}{(N_I - 1) * \frac{\mu_I^2}{N_I} + \sigma_I^2}$$

N_I—number of convolutional features
σ_I—standard deviation of convolutional features

Step 3: During decodification, the four-layer convolution inverse technique is used, and the size of the features is augmented using an up-sampling procedure that adds duplicates to the number of samples with convolutional features.

Step 4: Two layers related to dense and flatten the results are involved to attain the final reconstructed signal.

$L_{dense} = Wt^F * \frac{\partial L_{max}^l}{\partial t}$

where
Wt^F weight estimated with respect to index features
$\frac{\partial L_{max}^l}{\partial t}$ derivative for the maximum polling features
Estimated flatten layers with respect to whole dimension
$L_{flatten} = L_{dense}(:)$

Step 5: The optimization algorithm performed using Aquila optimization, with the addition to the stochastic gradient descent along with learning rate 0.001.

$\Delta Opt_L = -\rho \frac{v_l}{\sqrt{s_l + \tau}}$

v_l—exponential average of gradients of the features
s_l—xponential average of square of gradients of the features
ρ—Initial learning rate 0.001
τ—constant value

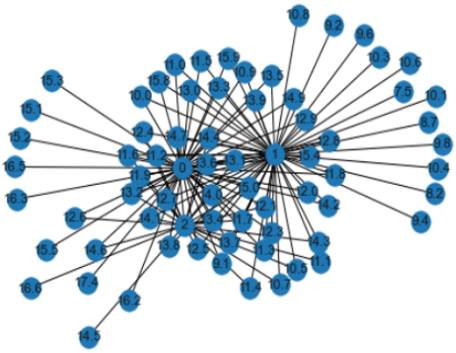
Step 6: After probability distribution, binary cross entropy is performed as loss function to estimate difference between actual samples

$loss = \sum_{i=1}^{k} B\,(Label^i) * \log\,(\Delta Opt_L^i)$

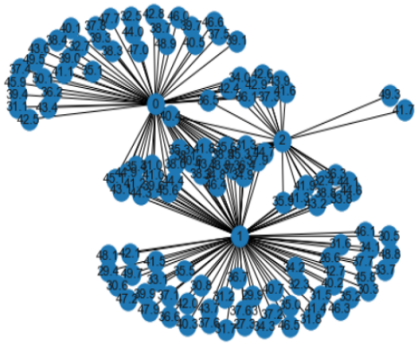
// $B(.)$ is the binary cross entropy operation

Step 7: The final classified output is,

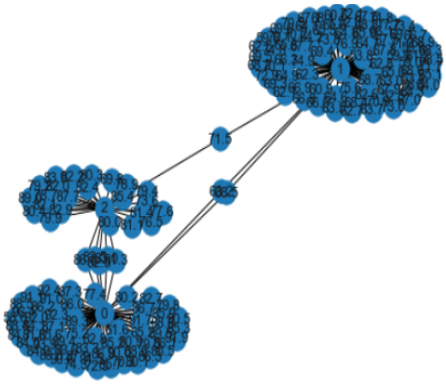
$Output_{cl} = Label^i(loss \leq 1)$



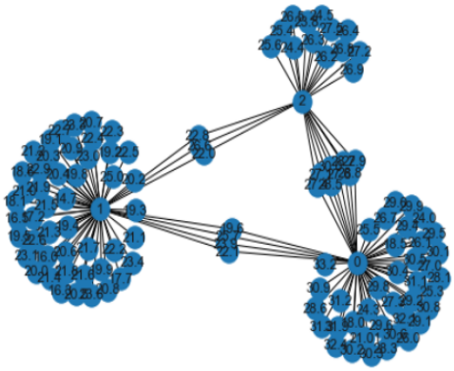
(a) Graph constructed for Hemoglobin (HB)



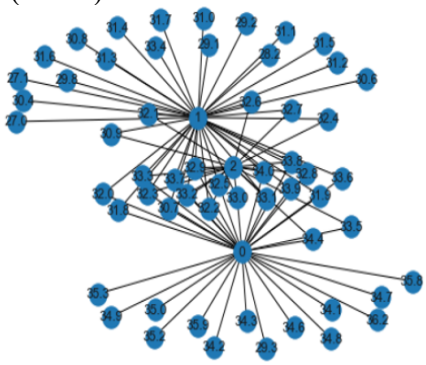
(b) Graph constructed for Hematocrit (HCT)



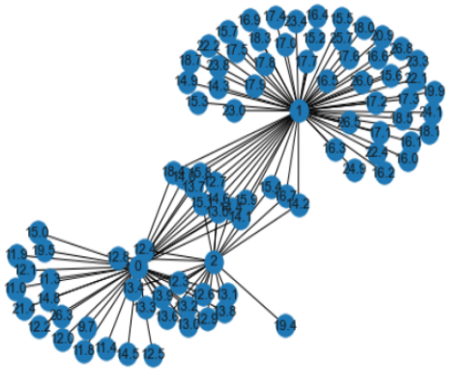
(c) Graph constructed for Mean corpuscular volume (MCV)



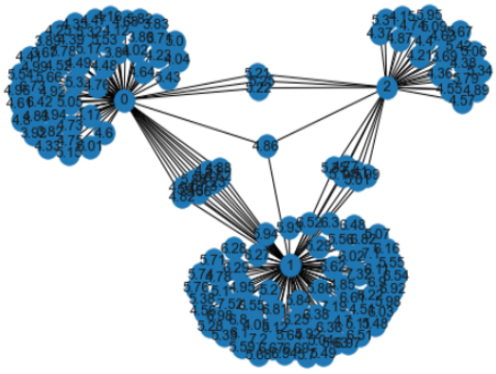
(d) Graph constructed for Mean corpuscular Hemoglobin (MCH)



(e) Graph constructed for Mean corpuscular Hemoglobin Concentration (MCHC)



(f) Graph constructed for RBC distribution width (RDW)



(g) Graph constructed for Red Blood Cell (RBC count)

Figure 2. Graph constructed for each attribute in Quantum graph theory model.

4. Performance analysis

The performance assessment of the proposed system is projected in this section. As a part of a collaborative research project done by Jawaharlal Nehru Technological University Institute, Hyderabad TEQIP-III, a memorandum of Understanding (MoU dated 15 Oct 2019) was executed among the Institute named Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering & Technology, Bachupally, Hyderabad, Telangana. The MoU aims at the objectives like data sharing and technology transfer. A dataset of 1387 patient's records were received who approached Thalassemia and Sickle cell Society (TSCS) for the purpose of diagnosis. Those records were shared as a MoU part and are pre-processed for maintaining the anonymity and confidentiality which satisfies patient's data privacy. Once it is pre-processed, few attributes were recognized with the class label at the blood sample diagnosis.

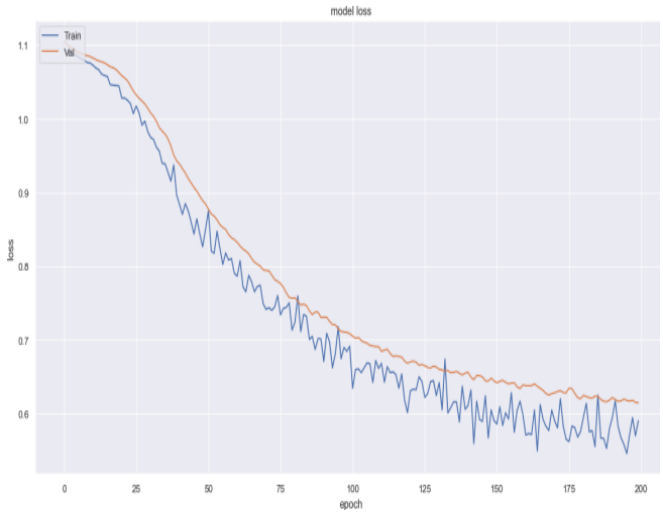


Figure 3. Model loss assessments for number of epochs.

Figure 3 signifies the assessment of model loss in terms of number of epochs for both testing and training data. As the number of epochs increased, the model loss decreases.

Figure 4. Comparison of attribute parameters range values (Low to high).

The attributes like Red Blood Cell (RBC) count, sex of the patient, Mean Corpuscular Volume (MCV), Mean Corpuscular Hemoglobin (MCH), RBC Distribution Width (RDW), Hematocrit (HCT), Mean Corpuscular Hemoglobin Concentration (MCHC), Hemoglobin (HB), and age and their corresponding range values low or high is represented for entire attributes. The rose color signifies higher value and blue denotes lower range. Based on the parameters range values variation, the disorder occurrence and cause are predicted.

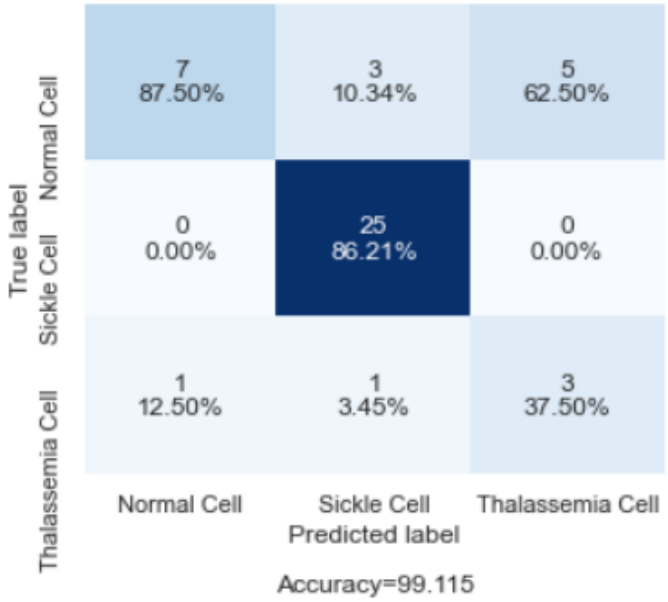


Figure 5. Confusion matrix representations for the accuracy assessment.

Figure 5 is the representation of confusion matrix formation and the accuracy assessment for the true label and predicted label values indicating normal cells, sickle cell, and Thalassemia cell. The confusion matrix is predicted for true label and predicted label. The accuracy assessment based on the constructed confusion matrix is about 99.115%. The confusion matrix signifies the count from the predicted and the actual values. The output TN denotes true negative values which signifies the number of negative samples that are classified accurately. Likewise, TP signifies the true positive values which signifies the number of positive samples that are classified accurately. FP indicates the values of false positive that is the actual number of negative samples that are classified as positive and FN indicates the value at which the actual positive samples are classified as negative.

Figure 6 is the comparative analysis made for the performance metrics like sensitivity, specificity, AUC, precision, and accuracy in terms of both (without quantum graph and with Quantum graph). The results indicate that the outcome attained with Quantum graph theory is having higher outcome values on comparing the ones without Quantum graph utilization. The application of Quantum graph model offers better result this is due to the advantage that Quantum algorithms for graph analysis is more effective on comparing their classical counterparts. In the condition of quantization that Eigen function on the graph cloud be defined over the associated plane-wave co-efficient. Since the Eigen function is stationary at Quantum evolution. For these benefits, the outcomes are better on employing Quantum graph theory model.

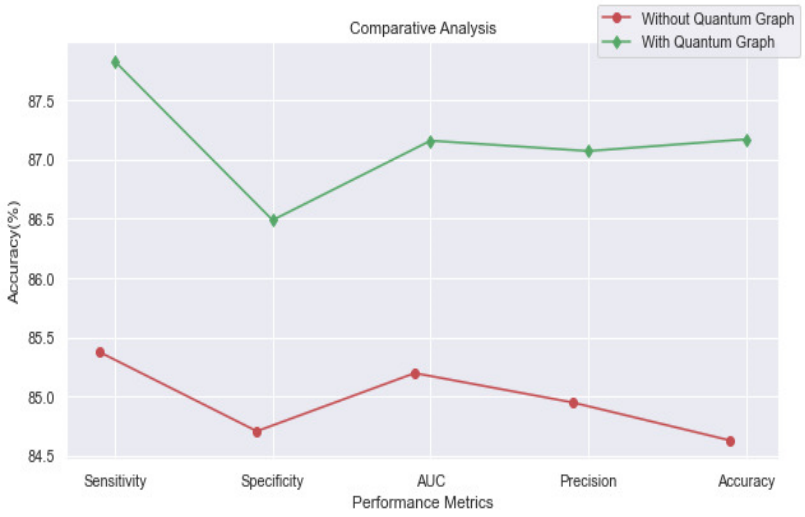


Figure 6. Comparative analysis of performance metrics attained for both (without Quantum graph and with Quantum graph).

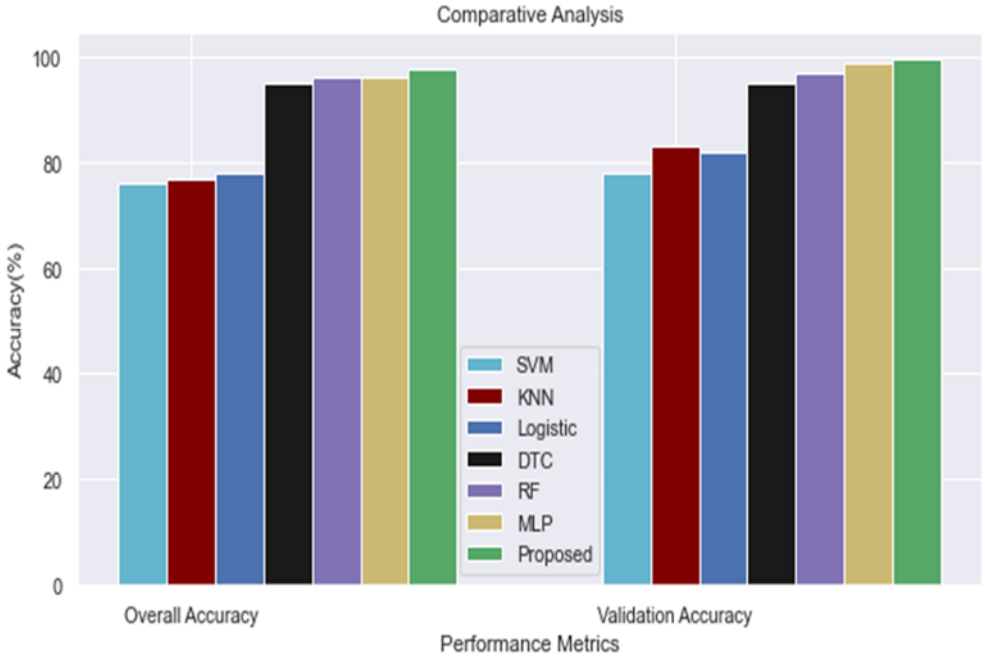


Figure 7. Comparative analysis of overall and validation accuracy [31].

The comparative estimation is made in Figure 7 for overall accuracy and validation accuracy for proposed classifier and existing classifiers like SVM, KNN, Logistic, DTC, RF, and MLP. From the outcome it was evident that proposed system is having higher accuracy rate than the traditional methods. Also, the suggested Quantum graph theory model with the Aquila optimization based cascaded LSTM is offering enhanced output than existing ones. This is due to the reason that Aquila optimization-based technique has advantages like simplicity, flexibility, and ability for avoiding local optima and meta-heuristic optimization approach. Also, this is capable of solving several complex and tricky optimization issues.

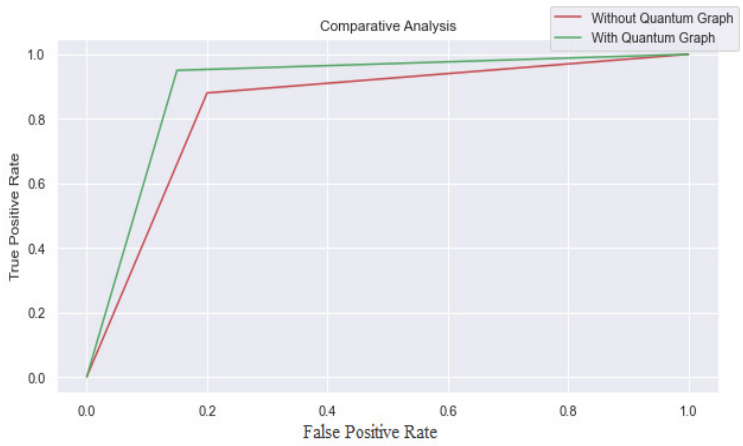


Figure 8. Comparative analysis of False and True Positive Rates for both (without Quantum graph and with Quantum graph).

Figure 8 is the comparative analysis made for the False and True Positive Rates in terms of both (without Quantum graph and with Quantum graph). The results indicate that the outcome attained with Quantum graph theory is having higher outcome values on comparing the ones without Quantum graph utilization.

Table 1. Performance comparison estimation of accuracy.

Methods	Accuracy (%)
Support Vector Machine (SVM)	76
K-Nearest neighbor (KNN)	77
Logistic Regression (L)	78
Decision Tree Classification (DTC)	95
Random Forest	96
Proposed	99.12

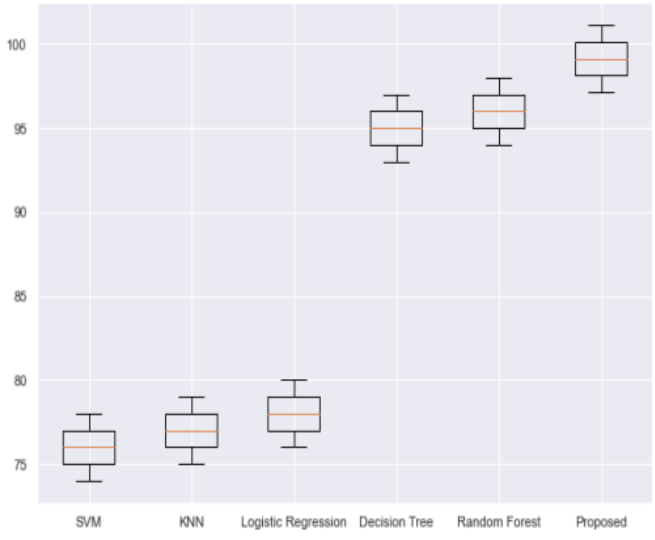


Figure 9. Performance comparison estimation of accuracy.

The comparative estimation is made in Table 1 in terms of accuracy for proposed classifier and existing classifiers like SVM, KNN, Logistic, DTC, and RF. From the outcome it was evident that proposed system is having higher accuracy rate than the traditional methods.

Figure 9 is the graphical representation of comparative estimation made in terms of accuracy for proposed classifier and existing classifiers like SVM, KNN, Logistic, DTC, and RF. From the outcome it was evident that proposed system classifier is having higher accuracy rate than the existing approaches. This is because the cascaded LSTM proposed is a generative approach which leverages LSTM approach for the prediction of topological infrastructures over the time of information cascade tree. It is competent of predicting the information cascades of various spatio-temporal growth patterns. For this reason, the presented cascaded LSTM is said to be effective.

5. Conclusions

The proposed work bids on presenting a mathematical modelling to identify monogenic disorder termed sickle cell anaemia with the use of classifier. This predictive model is effective in offering high accuracy rate of prediction process. Firstly, input DNA sequence is taken and the elastic property features were extracted using Quantum Graph theory model. Then, the extracted properties were optimized using Aquila optimization and classified using cascaded LSTM for attaining classified outcome of sickle cell and normal cells. In conclusion, the performance assessment was carried and the outcomes attained shows improvement in terms of sensitivity (3%), specificity (2.2%), and AUC (1.9%), precision (2%), and accuracy (2.6%) improvement was achieved on applying Quantum graph theory on comparing without Quantum graph theory application. The accuracy was compared with existing classifier like SVM, KNN, Logistic Regression, DTC, RF and MLP algorithms to validate the proposed system effectiveness. The simulations results illustrates that the presented classification offers 99.12% accuracy for predicting and classifying normal and sickle cell anaemia. As of the outcome attained it was evident that proposed system is having higher accuracy rate on comparing traditional approaches. In future, this work can be extended by predicting various other diseases and the algorithm can be enhanced so as to attain high performance.

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. M. C. V. Schie, S. Jainandunsing, J. E. R. Lennep, Monogenetic disorders of the cholesterol metabolism and premature cardiovascular disease, *Eur. J. Pharmacol.*, **816** (2017), 146–153. https://doi.org/10.1016/j.ejphar.2017.09.046
2. J. Zhou, Q. Lu, R. Xu, L. Gui, H. Wang, EL_LSTM: prediction of DNA-binding residue from protein sequence by combining long short-term memory and ensemble learning, *IEEE/ACM Trans. Comput. Biol. Bioinf.*, **17** (2018), 124–135. https://doi.org/10.1109/TCBB.2018.2858806
3. Q. Zhang, P. Liu, X. Wang, Y. Zhang, Y. Han, B. Yu, StackPDB: predicting DNA-binding proteins based on XGB-RFE feature optimization and stacked ensemble classifier, *Appl. Soft Comput.*, **99** (2021), 106921. https://doi.org/10.1016/j.asoc.2020.106921

4. S. Demirci, N. Uchida, J. F. Tisdale, Gene therapy for sickle cell disease: An update, *Cytotherapy*, **20** (2018), 899–910. https://doi.org/10.1016/j.jcyt.2018.04.003
5. V. M. Pinto, M. Balocco, S. Quintino, G. L. Forni, Sickle cell disease: A review for the internist, *Int. Emerg. Med.*, **14** (2019), 1051–1064. https://doi.org/10.1007/s11739-019-02160-x
6. X. Yang, Q. Zhou, W. Zhou, M. Zhong, X. Guo, X. Wang, et al., A cell-free DNA barcode-enabled single-molecule test for noninvasive prenatal diagnosis of monogenic disorders: Application to β-thalassemia, *Adv. Sci.*, **6** (2019), 1802332. https://doi.org/10.1002/advs.201802332
7. C. M. Dasari, R. Bhukya, Explainable deep neural networks for novel viral genome prediction, *Appl. Intell.*, **52** (2022), 3002–3017. https://doi.org/10.1007/s10489-021-02572-3
8. J. T. Shieh, M. Penon-Portmann, K. H. Wong, M. Levy-Sakin, M. Verghese, A. Slavotinek, et al., Application of full-genome analysis to diagnose rare monogenic disorders, *NPJ Genomic Med.*, **6** (2021), 1–10. https://doi.org/10.1038/s41525-021-00241-5
9. S. Mettananda, D. R. Higgs, Molecular basis and genetic modifiers of thalassemia, *Hematol. Oncol. Clin.*, **32** (2018), 177–191. https://doi.org/10.1016/j.hoc.2017.11.003
10. B. Chakraborty, Genetic algorithm with fuzzy fitness function for feature selection. In *Industrial Electronics*, *2002. ISIE 2002. Proceedings of the 2002 IEEE International Symposium on*, **1** (2002), 315–319. https://doi.org/10.1109/ISIE.2002.1026085
11. D. Singh, S. P. Singh, Self-organization and learning methods in short term electric load forecasting: A review, *Electr. Power Compon. Syst.*, **30** (2002), 1075–1089. https://doi.org/10.1080/15325000290085370
12. E. C. Morabito, M. Versaci, A fuzzy neural approach to localizing holes in conducting plates, *IEEE Trans. Magn.*, **37** (2001), 3534–3537. https://doi.org/10.1109/20.952655
13. L. Abualigah, A. Diabat, S. Mirjalili, M. Abd Elaziz, A. H. Gandomi, The arithmetic optimization algorithm, *Comput. Methods Appl. Mech. Eng.*, **376** (2021), 113609. https://doi.org/10.1016/j.cma.2020.113609
14. J. O. Agushaka, A. E. Ezugwu, L. Abualigah, Dwarf mongoose optimization algorithm, *Comput. Methods Appl. Mech. Eng.*, **391** (2022), 114570. https://doi.org/10.1016/j.cma.2022.114570
15. L. Abualigah, M. Abd Elaziz, P. Sumari, Z. W. Geem, A. H. Gandomi, Reptile Search Algorithm (RSA): A nature-inspired meta-heuristic optimizer, *Exp. Syst. Appl.*, **191** (2022), 116158. https://doi.org/10.1016/j.eswa.2021.116158
16. O. N. Oyelade, A. E. S. Ezugwu, T. I. Mohamed, L. Abualigah, Ebola optimization search algorithm: A new nature-inspired metaheuristic optimization algorithm, *IEEE Access*, **10** (2022), 16150–16177. https://doi.org/10.1109/ACCESS.2022.3147821
17. L. Abualigah, D. Yousri, M. Abd Elaziz, A. A. Ewees, M. A. Al-Qaness, A. H. Gandomi, Aquila optimizer: a novel meta-heuristic optimization algorithm, *Comput. Ind. Eng.*, **157** (2021), 107250. https://doi.org/10.1016/j.cie.2021.107250
18. J. A. López-Rivera, E. Pérez-Palma, J. Symonds, A. S. Lindy, D. A. McKnight, C. Leu, et al., A catalogue of new incidence estimates of monogenic neurodevelopmental disorders caused by de novo variants, *Brain*, **143** (2020), 1099–1105. https://doi.org/10.1093/brain/awaa051
19. E. K. L. Chiu, W. W. I. Hui, R. W. K. Chiu, cfDNA screening and diagnosis of monogenic disorders–where are we heading, *Prenatal Diagn.*, **38** (2018), 52–58. https://doi.org/10.1002/pd.5207

20. A. M. Quinn, B. N. Valcarcel, M. M. Makhamreh, H. B. Al-Kouatly, S. I. Berger, A systematic review of monogenic etiologies of nonimmune hydrops fetalis, *Genet. Med.*, **23** (2021), 3–12. https://doi.org/10.1038/s41436-020-00967-0
21. M. E. Niemi, H. C. Martin, D. L. Rice, G. Gallone, S. Gordon, M. Kelemen, et al., Common genetic variants contribute to risk of rare severe neurodevelopmental disorders, *Nature*, **562** (2018), 268–271. https://doi.org/10.1038/s41586-018-0566-4
22. M. Zech, R. Jech, S. Boesch, M. Škorvánek, S. Weber, M. Wagner, et al., Monogenic variants in dystonia: an exome-wide sequencing study, *Lancet Neurol.*, **19** (2020), 908–918. https://doi.org/10.1016/S1474-4422(20)30312-4
23. D. M. Connaughton, C. Kennedy, S. Shril, N. Mann, S. L. Murray, P. A. Williams, et al., Monogenic causes of chronic kidney disease in adults, *Kidney Int.*, **95** (2019), 914–928. https://doi.org/10.1016/j.kint.2018.10.031
24. G. Valles-Ibáñez, A. Esteve-Sole, M. Piquer, E. A. González-Navarro, J. Hernandez-Rodriguez, H. Laayouni, et al., Evaluating the genetics of common variable immunodeficiency: monogenetic model and beyond, *Front. Immunol.*, **9** (2018), 636. https://doi.org/10.3389/fimmu.2018.00636
25. J. M. Alperin, L. Ortiz-Fernández, A. H. Sawalha, Monogenic lupus: a developing paradigm of disease, *Front. Immunol.*, **9** (2018), 2496. https://doi.org/10.3389/fimmu.2018.02496
26. J. J. Ashton, E. Mossotto, I. S. Stafford, R. Haggarty, T. A. Coelho, A. Batra, et al., Genetic sequencing of pediatric patients identifies mutations in monogenic inflammatory bowel disease genes that translate to distinct clinical phenotypes, *Clin. Transl. Gastroenterol.*, **11** (2020). https://doi.org/10.14309/ctg.0000000000000129
27. J. C. Almlöf, S. Nystedt, D. Leonard, M. L. Eloranta, G. Grosso, C. Sjöwall, et al., Whole-genome sequencing identifies complex contributions to genetic risk by variants in genes causing monogenic systemic lupus erythematosus, *Hum. Genet.*, **138** (2019), 141–150. https://doi.org/10.1007/s00439-018-01966-7
28. S. Vidal, N. Brandi, P. Pacheco, J. Maynou, G. Fernandez, C. Xiol, et al., The most recurrent monogenic disorders that overlap with the phenotype of Rett syndrome, *Eur. J. Paediatr. Neurol.*, **23** (2019), 609–620. https://doi.org/10.1016/j.ejpn.2019.04.006
29. S. Butscheidt, A. Delsmann, T. Rolvien, F. Barvencik, M. Al-Bughaili, S. Mundlos, et al., Mutational analysis uncovers monogenic bone disorders in women with pregnancy-associated osteoporosis: three novel mutations in LRP5, COL1A1, and COL1A2, *Osteoporosis Int.*, **29** (2018), 1643–1651. https://doi.org/10.1007/s00198-018-4499-4
30. F. Cerrato, A. Sparago, F. Ariani, F. Brugnoletti, L. Calzari, F. Coppedè, et al., DNA methylation in the diagnosis of monogenic diseases, *Genes*, **11** (2020), 355. https://doi.org/10.3390/genes11040355
31. S. Yeruva, M. S. Varalakshmi, B. P. Gowtham, Y. H. Chandana, P. K. Prasad, Identification of sickle cell anemia using deep neural networks, *Emerging Sci. J.*, **5** (2021), 200–210. https://doi.org/10.28991/esj-2021-01270

AIMS Press