[1]Mojtaba aliabadian

[2*]Ali A. Safaei

[3]Mehdi Hosseinzadeh

Development a Hybrid Model based Deep Learning to Diagnosis of Autoimmune Diseases

Journal of Electrical Systems

Abstract: - Systemic Lupus Erythematosus (SLE) is a chronic autoimmune disease that can cause pimples and butterfly-shaped hives on the skin of the nose and cheeks, and if left untreated, can spread to the entire body. Apart from damaging the skin, lupus can cause inflammation or damage to the joints, muscles, inner membrane or around the lung, heart, kidney and brain. Considering the importance of this disease and the early diagnosis of ANF, in this study, an approach based on deep learning and gray wolf meta-heuristic algorithm was discussed in order to identify this disease among patients. In this study is from Taiwan Precision Medicine Initiative (TPMI), which was used by 946 patients with their own diseases in the period of June 2019 to June 2020. The results showed that the use of the proposed method can increase the detection accuracy to 0.8936. Meanwhile, for LR, RF, SVM, LGBM, GBT, and XGB models, the detection accuracy is 0.7887, 0.8345, 0.7729, 0.7993, 0.7975, and 0.8345, respectively.

Keywords: Autoimmune Diseases, Systemic lupus erythematosus (SLE), Lupus, Deep Learning, Machine Learning

I. INTRODUCTION

In today's medical world, one of the diseases that are caused by the combination of many factors and are considered chronic diseases, are diseases with autoimmune origin. Accumulation of autoimmune complexes can cause systemic and local inflammation. In some patients with chronic and autoimmune diseases, the deposition of autoimmune complexes and subsequent inflammation in tissues leads to irreversible organ damage and disease burden [1]. In general, autoimmune diseases account for at least 7% of the diseases of the United States population and create a significant socio-economic burden on the health status. Although extensive research in recent years has identified several genetic factors related to autoimmune diseases, current research has not yet been able to adequately determine the cause of the disease, the age of onset, and the progression of autoimmune diseases [2]. To date, more than 80 types of autoimmune diseases have been identified. Previous studies estimate that they collectively affect 3-9% of the world's population, with this number increasing with age. Autoimmune diseases are more common among women than men, so the World Health Organization (WHO) has estimated the ratio of autoimmune diseases between women and men to be around 2:4 to 2:7. However, estimates vary widely based on the specific disease and population studied. Some autoimmune diseases, such as systemic lupus erythematosus, may worsen during pregnancy, while the symptoms of other diseases, such as rheumatoid arthritis, may improve. In addition, autoimmune diseases may also affect children's long-term health and increase the risk of developing neurodevelopmental disorders such as autism. A 2011 Danish National Hospital Register study evaluated one million women aged 14-32 for 12 years and found that 2.4% had an autoimmune disease [3].

Systemic lupus erythematosus (SLE) is a chronic inflammatory disease that affects multiple organs and is characterized by a wide variety of autoantibodies. Approximately half of patients with SLE develop renal disease (usually immune complex-mediated glomerular disease), and kidney damage is a major cause of morbidity and mortality [4,5]. According to studies and researches, SLE usually develops between teenagers and adults, ranging from 15 to 44 years of age, this disease is characterized by chronic pain and inflammation in multiple systems including kidneys, brain, blood cells, joints, skin. , lungs and heart, and a series of symptoms such as fatigue, fever, joint problems, confusion and memory loss, and skin lesions, which often appear in the form of butterflies on the nose and cheeks, show themselves. These symptoms can develop suddenly or slowly, depending on the specific cause and which body systems are affected [6]. The presence of haematuria, proteinuria or decreased kidney function raises suspicion for lupus nephritis, which is confirmed by kidney biopsy [7]. The 2003 classification system of the International Society of Pathology of Nephrology (ISN/RPS) divides kidney pathology in lupus nephritis into six categories: Class I includes mesangial immune deposits with mesangial hypercellulitis. Class II consists of mesangial immune deposits with mesangial hypercellulitis. Class III includes proliferative glomerulonephritis involving less than 50% of glomeruli with active and/or chronic lesions. Class

[1] [1] *Department Of Computer Engineering,Qeshm Branch, Islamic Azad University, Qeshm, Iran*

2 *Department Of Medical Informatics,Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran*

3 *Institute of Research and Development , Duy Tan University , Da Nang , Vietnam*

* Corresponding author

Copyright © JES 2024 on-line : journal.esrgroups.org

IV, proliferative glomerulonephritis includes more than 50% of glomeruli with active and/or chronic lesions. Class V, as membranous nephritis, and Class VI, includes advanced sclerosing lesions. Alternatively, class III, IV, and V cells can coexist and morph between histologic patterns. Also, proliferative lupus nephritis (PLN) is described as lupus nephritis with or without class V nephropathy. In contrast, membranous lupus nephritis (MLN) is defined as lupus nephritis that has pure (class V) membranous nephropathy (MLN) [8]. Due to the fact that this disease manifests itself in various ways, it is often confused with other diseases, and unfortunately, its main cause, which can be an autoimmune disease, is not considered. In fact, each doctor prescribes medicine, suggests medical tests, or recommends hospitalization based on his personal thoughts and experiences. Therefore, gathering these experiences and skills and creating a knowledge base and, as a result, developing an intelligent diagnostic experience system that can diagnose this disease earlier and avoid huge costs and save the patient from confusion, seems necessary. . Because the global consensus is that chronic diseases are one of the main drivers of wasting health costs. When talking about the impact of chronic diseases on the global and national economy, generally both direct and indirect costs should be considered [9].

Therefore, in order to achieve better results and reduce the costs of care for patients, the need to implement and use preventive behaviors and measures (improve the health knowledge of patients), increase self-care skills and empower patients, use evidence-based care programs, improve performance of information system along with electronic (online) patient-centered programs in order to track and follow the treatment and progress of the disease, the cooperation of all stakeholders in the health care system, including patients, caregivers, financial sponsors, the general public, and the government in order to design and compile the prevention and care programs based on evidence are felt more and more throughout people's lives. This problem has caused the use of artificial intelligence-based approaches, as an effective tool in the development of medical services, to play a significant role in diagnosing this disease and providing care solutions. This problem has caused the use of artificial intelligence and deep learning techniques to become the focus of many researchers in order to diagnose autoimmune diseases such as SLE. For example, in [10] advanced magnetic resonance spectroscopy using a deep learning approach has been used to diagnose neuropsychiatric systemic lupus erythematosus (NPSLE). The accuracy of the proposed method was equal to 97.5%. In [11], the pile and propeller algorithm enhanced with support vector machine is used to detect SLE. The accuracy of this model is equal to 92.04%. In [12], different machine learning models have been used to detect electrocardiogram abnormalities in patients with SLE. Among the used algorithms, the random forest model has the highest accuracy of 94.85%. In [13], by applying the deep learning algorithm, the accuracy of 94.56% has been reached in order to diagnose the systemic lupus erythematosus disease using the national database of designated incurable diseases of Japan. In [14], six machine learning models namely logistic regression, random forest (RF), support vector machine, light gradient boosting machine, gradient tree boosting and extreme gradient boosting (XGB) have been used to identify SLE patients. RF and XGB models with 83.45% accuracy have the best performance among these models.

According to the investigations carried out in the literature review section, where we evaluated different methods of deep learning in the diagnosis of diseases, especially SLE, it was found that the use of new neural network methods, especially deep learning, is a new method in obtaining disease symptoms. On the other hand, considering the effective causes and known environmental factors in lupus disease, including genetic and epigenetic factors, sunlight, mental and physical stress, sex hormones, smoking, infectious agents such as viruses and consumption some drugs, the occurrence of this disease is observed more and more day by day. Due to the fact that so far no accurate and fast automatic allocation system has been designed in this field, therefore, providing an automatic system at the disposal of internal and general physicians or even available to the general public is of great importance and attention. Therefore, in the present study, an attempt is made to provide a model to predict lupus disease, an efficient model that has the appropriate speed and accuracy. The proposed approach is based on deep learning and specifically, the combination of LSTM algorithm and gray wolf algorithm (GWO).

II. MATERIAL AND METHOD

A) Dataset

In this study, a retrospective case-control design is used using data from the Taiwan Precision Medicine Initiative (TPMI) presented in [14]. This dataset consists of participants at Taichung Veterans General Hospital (TCVGH), Taiwan, from June 2019 to June 2020. The primary goal of TPMI was to incorporate genetic information into clinical applications. A blood sample was collected from each participant enrolled in TPMI, extracted for DNA, and genotyped. The genetic profile of TPMI participants is linked to their electronic health records for case management and precision medicine implementation. On the other hand, 946 patients were tested using ANA with a positive titer equal to or greater than 1:80 in order to measure the classification criteria of International Systemic Lupus Clinics (SLA) in them.

B) Methodology

The approach used in this research is based on deep learning. Deep learning is one of the most powerful methods in the field of artificial intelligence, and many researchers have focused on deep learning structures to improve computational power. However, deep learning architecture also has disadvantages. One of the most important disadvantages of this model is that deep learning does not facilitate a comprehensive theoretical understanding of learning. A well-known weakness of deep learning methods is their black-box nature. Deep learning models usually include non-linear functions. For this reason, not only there is no simple mapping from input to output, but also the effect of changing one input can be highly dependent on the values of other inputs. This makes it difficult to conceptually understand what is happening in deep neural networks. The more precise the model, the less interpretable it is. It means that complex algorithms such as deep learning are difficult to understand, but their accuracy is high. Therefore, due to the high accuracy of deep neural networks, they have been used for fault finding. In general, deep algorithms are divided into four categories: convolutional neural networks, closed Boltzmann machine, self-encoders and sparse coding. The use of closed Boltzmann machine methods is time-consuming and requires long calculations. On the other hand, in sparse coding, feature training is not possible. Among these methods, convolutional neural networks have more applications than other methods and it is used far more [15].

The flowchart of proposed method is shown in Fig1.

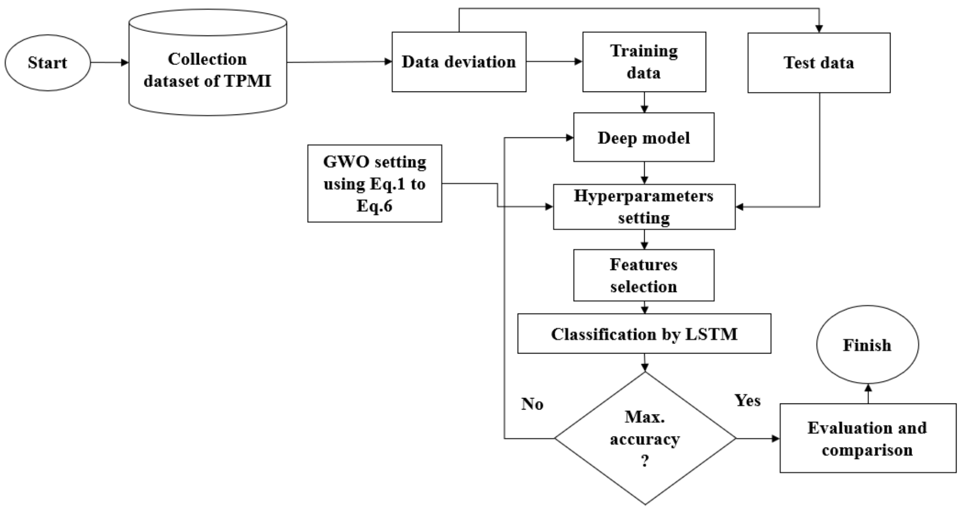

Fig1. The flowchart of proposed method

As shown in Fig.1, after reading the data set related to the clinical data of lupus patients and pre-processing on the data, the selection of optimal features is performed using the Gray Wolf Algorithm (GWO). The gray wolf optimization algorithm simulates the hunting stages of wolves. The structure of hunting includes three parts: stalking and encircling the prey, harassing the prey until it stops, and finally attacking the prey. Each wolf i as a solution to the problem in the search space has a position vector $W_i =< w_{i1}, w_{i2}, \dots, w_{in} >$ which represents the n dimensions of the problem. To evaluate the position of the wolves, the fitting function (suitable to the problem definition) is used. According to the values of the fitting function, the first best wolf is indicated by alpha (α), the second best wolf by beta (β) and the third best wolf by (δ). During the hunting (optimization) process, the wolves update their positions according to the positions of the three wolves, alpha, beta, and delta. At the end, the algorithm returns alpha wolf as the final solution. In this algorithm, N is the number of flocks and D is the number of decision variables or dimensions of the optimization problem. Therefore, the pack of gray wolves is simulated by an N*D matrix. Each row corresponds to a possible solution of the optimization problem. In the proposed model, N is the number of data set records and D is the number of features. The herd population, which includes a large number of wolves, is defined according to Eq. (1). In the proposed model, the working method for the electric energy data set is such that the gray wolf algorithm consists of N flocks and each flock is composed of a wolf (feature). Each herd is defined by these D features [17].

(1) $$Population\ of\ GWO = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{12} \\ x_{22} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nd} \end{bmatrix}$$

In the set $x_i = (x_{i1}, x_{i2}, \dots, x_{id}), i = 1,2, \dots, n$, each x_i represents a possible solution in the solution space. Each pack of wolves consists of a group of attacking wolves, which are considered as elements of a solution. All the attacking wolves in a wolf pack are considered as a whole unit, moving towards a suitable place with abundant resources. If the pack of wolves reaches an ideal position, an optimal solution is obtained. The evaluation of each wolf pack is calculated based on the objective function according to equation (2) [16].

$$fit_i = 1 - \frac{Obj_i - worst(Obj)}{best(Obj) - worst(Obj)} \quad (2)$$

In equation (2), fit_i is the fitness of the ith herd. The parameter Obj_i is the value of the objective function for the i-th herd. Each pack of wolves is calculated based on distance criteria. Worst and best parameters are the worst and the best wolf pack relative to the prey. In the proposed model, the gray wolves algorithm should be changed from continuous to discrete. To convert the numbers into binary, the v-shaped hyperbolic tangent function of two solutions according to equation (3) is used. In the proposed method, which is due to random stepping, they have continuous values and therefore must be converted into a binary solution.

$$y^k = |tanhx^k| \quad)4($$

$$x_{ij} = \begin{cases} 0, & if\ rand < y^k \\ 1, & otherwise \end{cases} \quad)5($$

In the proposed model, a subset of features that lead to the most optimal value is selected using the gray wolves algorithm. The fitness function for feature selection from each wolf pack is defined according to equation (6). In equation (6), |n| is the total number of features and |S| is the number of selected features. The accuracy parameter is the percentage of accuracy and the value of the parameters δ and ρ are fixed and their values are equal to 1 and 99, respectively.

$$Fitness = \delta.Accuracy + \rho.\frac{|n| - |S|}{|n|} \quad)6($$

In the next step, LSTM algorithm is used to classify features. LSTM is a recurrent neural network that has the ability to learn long dependencies, so it can process all data sequences, and because it has short-term memory, there is no fading problem for long expressions. An LSTM unit consists of a cell and a gate. The input is a forget gate and an output gate. The structure of LSTM is similar to recurrent neural networks, except that instead of each hidden unit connected to itself, there is some kind of memory in them. Although recurrent neural networks are designed to work with sequences of different lengths, they have limitations in this field, which makes recurrent neural networks unable to learn long dependencies. The LSTM network has solved this problem by using memory units. This model includes three layers of Dense LSTM embedding. The task of the first layer, embedding, is to convert words into 32 vectors with the structure of real numbers. The maximum number of words in this layer is 10,000. The second layer is an LSTM layer made of 32 neurons. The final layer is the desnse layer, which has the ability to recognize three classes, positive, negative and neutral. In this model, the Softmax function is used as an activation function to map the output of the model between zero and 1, and to train the model, the Adam optimizer function is used [17].

In figure (2), the schematic of the proposed deep layers model is presented.

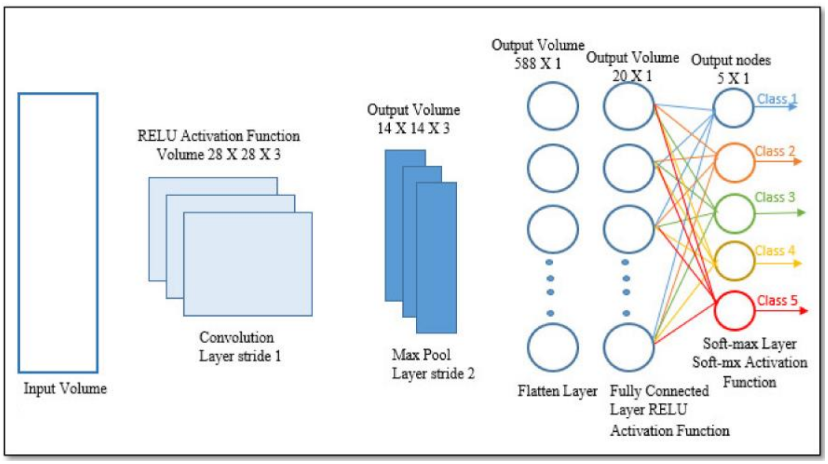

Fig2. Architecture of deep network layers

At finally, in order to evaluate the performance of the proposed model, the confusion matrix based on the following equations is used.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Specificity = \frac{TN}{TN + FP} \quad (9)$$

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

$$F1 - Score = 2 \times (Precision \times Recall)/(Precision + Recall) \quad (11)$$

III. Results

By applying the GWO algorithm and adjusting the parameters of the model, among the 131 main features, finally 114 features are selected and the values of these features are considered as the input features of the LSTM model. The convergence diagram of the GWO algorithm is presented in Figure (3) and the results of applying this algorithm are presented in Table (1).

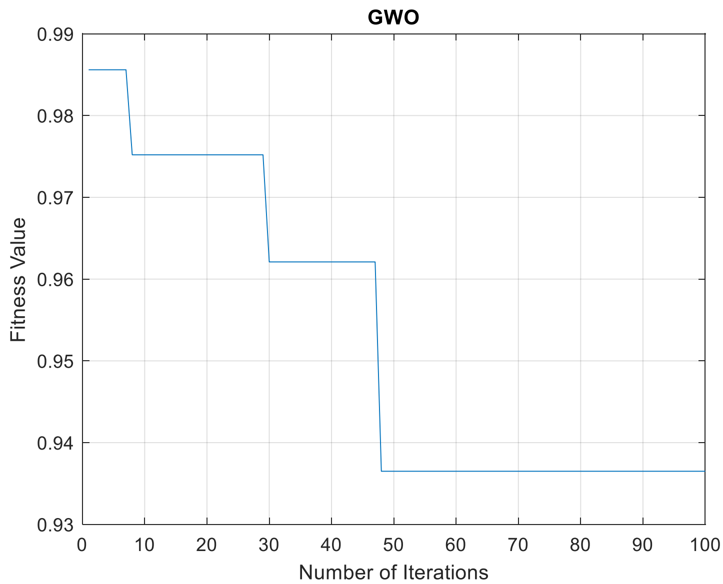

Fig3. GWO algorithm convergence diagram for feature selection

Table 1: The result of applying the GWO algorithm

Wolves No.	No. of Original Features	No. of Selected Features by ABC	Accuracy (%)
30	131	114	**93.65**

The settings of this algorithm are given in Table 2:

Table 2: Initial parameters setting of LSTM

Parameters	Value
Data in training process	80 %
Data in testing process	20 %
No. of hidden layers	200
No. of epoch	400

Training function	Adam
Gradient threshold	1
Initial learning rate	0.005
Coefficient of learning rate drop	0.3

By implementing the proposed model on the data set, the prediction results have been presented. Figure 3 shows the convergence diagram of the LSTM model, which reached an RMSE error of 0.216 after 400 epochs.

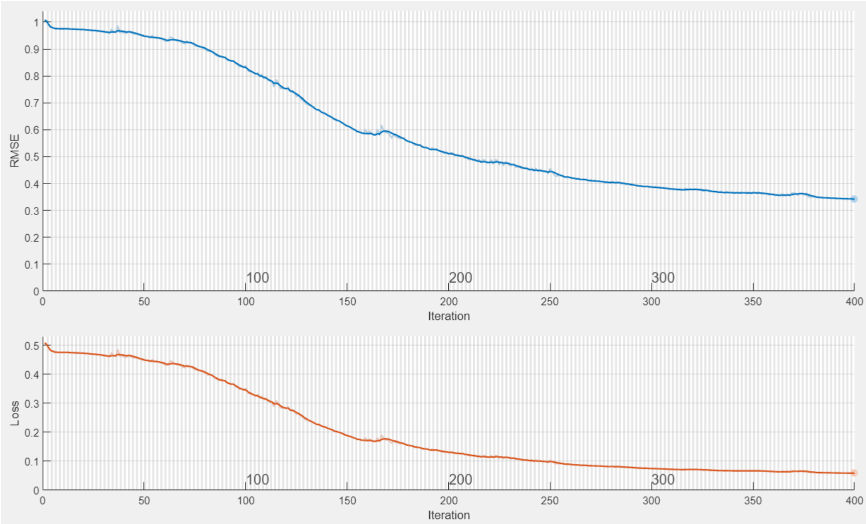

Fig4. Convergence diagram of LSTM model

In Table (3) the results obtained from the application of proposed algorithm in estimating the status of lupus disease are presented in comparison of other ML models in [14].

Table (3): Model performance comparison using TPMI testing dataset

Model	Accuracy	Precision	Specificity	F1 Score	Ref.
LSTM	0.8936	0.8359	0.9221	0.8831	Present Study
LR	0.7887	0.6949	0.8575	0.6721	[14]
RF	0.8345	0.7746	0.8971	0.7403	[14]
SVM	0.7729	0.6429	0.8021	0.6767	[14]
LGBM	0.7993	0.7193	0.8734	0.6833	[14]
GBT	0.7975	0.7033	0.8575	0.6900	[14]
XGB	0.8345	0.7684	0.8918	0.7432	[14]

In order to better understand the results of the above table, the diagram of Figure (4) is presented. As can be seen, in all evaluation criteria, the proposed method has a better performance.

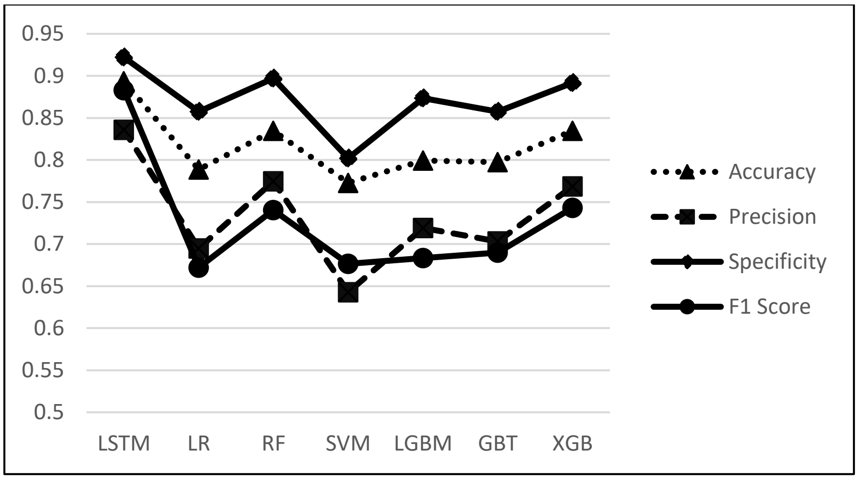

Fig5. Comparing evaluation criteria of models

IV. Conclusion

Lupus is one of the types of autoimmune diseases that result from the attack of the body's immune system on its own tissues and organs. These attacks cause inflammation, swelling and damage to different parts of the body such as joints, skin, kidneys, blood, heart and lungs. Although there is no complete cure for lupus, there are treatments to control the disease and reduce its severity. In recent years, with the increase of expert systems and with the significant progress of artificial intelligence and machine learning, the efforts to study and research on the issue of making organizational systems intelligent have increased and researchers have achieved good results in this field. The increase of machine learning algorithms as well as the ever-increasing development of these algorithms, many researchers have started investigating and studying these capabilities and are taking steps towards the progress of organizations. Considering the importance of this type of disease all over the world, especially in Iran, in this research, an attempt was made to provide an automatic system for the early detection and diagnosis of lupus by using artificial intelligence techniques. Accordingly, after data preparation, the integrated gray wolf approach was used as the feature selection algorithm in combination with the LSTM deep network algorithm. The data used in this study is from Taiwan Precision Medicine Initiative (TPMI), which was used by 946 patients with their own diseases in the period of June 2019 to June 2020. Based on this, in the first step after data preparation, the gray wolf algorithm was used to select the optimal features. This algorithm can select 114 features out of 131 primary features for LSTM model inputs with 93.46% accuracy. In the next step, by adjusting the hyperparameters of the LSTM model and dividing the data (80% of the training data and 20% of the test data), data classification was done. The results showed that the LSTM model reached an RMSE error of 0.216 after 400 iterations. Finally, the performance of the proposed model was compared with late machine learning methods. The results showed that the use of the proposed method can increase the detection accuracy to 0.8936. Meanwhile, for LR, RF, SVM, LGBM, GBT, and XGB models, the detection accuracy is 0.7887, 0.8345, 0.7729, 0.7993, 0.7975, and 0.8345, respectively. Also, the F1-score for the proposed model is equal to 0.8831 and for LR, RF, SVM, LGBM, GBT and XGB models it is equal to 0.6721, 0.7403, 0.6767, 0.6833, 0.6900 and 0.7432 respectively.

V. References

[1] Ramaswamy, M., Tummala, R., Streicher, K., Nogueira da Costa, A., & Brohawn, P. Z. (2021). The pathogenesis, molecular mechanisms, and therapeutic potential of the interferon pathway in systemic lupus erythematosus and other autoimmune diseases. International Journal of Molecular Sciences, 22(20), 11286.

[2] Gulati, G., & Brunner, H. I. (2018, April). Environmental triggers in systemic lupus erythematosus. In Seminars in arthritis and rheumatism (Vol. 47, No. 5, pp. 710-717). WB Saunders.

[3] Nielsen, T. C., Nassar, N., Harrison, C., Shand, A., Dale, R. C., Lowe, S., & Lain, S. J. (2022). Prevalence of autoimmune disease among pregnant women and women of reproductive age in New South Wales, Australia: a population-based study. The Journal of Maternal-Fetal & Neonatal Medicine, 35(17), 3229-3237.

[4] Wang, M., Liang, Y., Hu, Z., Chen, S., Shi, B., Heidari, A. A., ... & Chen, X. (2022). Lupus nephritis diagnosis using enhanced moth flame algorithm with support vector machines. Computers in Biology and Medicine, 145, 105435.

[5] Parikh, S. V., Almaani, S., Brodsky, S., & Rovin, B. H. (2020). Update on lupus nephritis: core curriculum 2020. American Journal of Kidney Diseases, 76(2), 265-281.

[6] Kinsey, D., Paul, C. P., Taylor, D., Caricchio, R., Kulathinal, R. J., & Hayes-Conroy, A. (2018). The whole lupus: Articulating biosocial interplay in systemic lupus erythematosus epidemiology and population disparities. Health & place, 51, 182-188.

[7] Parikh, S. V., Almaani, S., Brodsky, S., & Rovin, B. H. (2020). Update on lupus nephritis: core curriculum 2020. American Journal of Kidney Diseases, 76(2), 265-281.

[8] Abbasifard, M., Mohammadiranjbar, F., Mohammad-Sadeghipour, M., Mahmoodi, M., Hassanshahi, G., Swann, J., ... & Hajizadeh, M. R. (2022). The Evaluation of the epigenetics related genes expression (DNMT, HDAC1) in Iranian patients with Systemic Lupus Erythematosus.

[9] Imran, T. F., Yick, F., Verma, S., Estiverne, C., Ogbonnaya-Odor, C., Thiruvarudsothy, S., ... & Kothari, N. (2016). Lupus nephritis: an update. Clinical and experimental nephrology, 20(1), 1-13.

[10] Li, X., Bai, L., Ge, Z., Lin, Z., Yang, X., & Zhou, T. (2021). Early diagnosis of neuropsychiatric systemic lupus erythematosus by deep learning enhanced magnetic resonance spectroscopy. *Journal of Medical Imaging and Health Informatics*, *11*(5), 1341-1347.

[11] Wang, M., Liang, Y., Hu, Z., Chen, S., Shi, B., Heidari, A. A., ... & Chen, X. (2022). Lupus nephritis diagnosis using enhanced moth flame algorithm with support vector machines. *Computers in Biology and Medicine*, *145*, 105435.

[12] Hu, Z., Wu, L., Lin, Z., Liu, X., Zhao, C., & Wu, Z. (2023). Prevalence and associated factors of Electrocardiogram abnormalities in patients with systemic lupus erythematosus: a machine learning study. *Arthritis Care & Research*, *74*(10), 1640-1648.

[13] Kimura, T., Ikeuchi, H., Yoshino, M., Sakate, R., Maruyama, S., Narita, I., & Hiromura, K. (2023). Profiling of kidney involvement in systemic lupus erythematosus by deep learning using the National Database of Designated Incurable Diseases of Japan. *Clinical and Experimental Nephrology*, *27*(6), 519-527.

[14] Chung, C. W., Chou, S. C., Hsiao, T. H., Zhang, G. J., Chung, Y. F., & Chen, Y. M. (2024). Machine learning approaches to identify systemic lupus erythematosus in anti-nuclear antibody-positive patients using genomic data and electronic health records. *BioData Mining*, *17*(1), 1.

[15] Javaid, A., Niyaz, Q., Sun, W., & Alam, M. (2016). A deep learning approach for network intrusion detection system. Eai Endorsed Transactions on Security and Safety, 3(9), e2.

[16] Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey wolf optimizer. *Advances in engineering software*, *69*, 46-61.

[17] Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena*, *404*, 132306.