NOVA MEDICA INFORMATICA

(Article)

Deep Learning-Based Diagnostic Support for Complex Autoimmune Syndromes: A Multi-label Classification Study

Jacky Tuamelly[1]*; Joestiantho L Kilmanun[2]

Received: 27-02-2026
Revised: 03-03-2026
Accepted: 03-03-2026
Published: 27-03-2026

Correspondence:
tuamellyjacky@gmail.com

[1] Pattimura University, Ambon, Indonesia; tuamellyjacky@gmail.com
[2] Pattimura University; Ambon, Indonesia; laurenzkilmanun@gmail.com

Highlights

The Multi-Label Multi-Layer Perceptron (ML-MLP) with 79 clinical parameters is capable of simultaneously recognizing co-existing autoimmune syndromes with high accuracy.

What are the main findings?

- A combination of 79 laboratory features, including both hematological and serological tests, achieved AUC values ranging from 0.98 to 1.00. This is significantly higher than the results of models using only single biomarkers.
- A Multi-Label Multi-Layer Perceptron (ML-MLP) approach is shown to be an efficient tool for recognizing the profile of polyautoimmunity. This allows for the simultaneous detection of co-existing pathological states in a single run.
- A stable learning convergence is shown with improvements in F1-score values ranging from 0.26 to 0.47.

What are the implications of the main findings?

- The application of this system in clinical settings will serve as a virtual assistant to aid physicians in efficiently integrating heterogeneous data, which will minimize diagnostic errors caused by cognitive fatigue.
- The study results reveal that the application of routine Complete Blood Count (CBC) and specialized serology will improve diagnostic accuracy without the need to rely exclusively on costly genomic tests.
- The study results will aid in the shift towards personalized medicine, where it is possible to diagnose overlap syndromes early, allowing for more comprehensive interventions in the management of complex immunologic disorders.

Abstract

Autoimmune conditions frequently show similar clinical signs and symptoms, making it difficult to accurately diagnose when there are coexisting conditions. This study proposes a computational framework that aims to simultaneously predict multiple autoimmune labels using routine clinical data and advanced serological tests. The study examined 13,812 patients with 79 diagnostic features, including hematologic tests, inflammatory tests, and specific autoantibodies. The study trained a multi-layer perceptron (MLP) to predict autoimmune conditions such as Systemic Lupus Erythematosus (SLE), Rheumatoid Arthritis, and Graves' Disease. The model achieved strong performance, with validation F1 scores starting at 0.4723 and near-perfect Area Under the Curve (AUC) values ranging from 0.98 to 1.00 for all categories. The 79-parameter approach was highly effective in describing the autoimmunity "mosaic" and identifying polyautoimmunity with high precision compared to contemporary models that use fewer autoantibodies. The study shows that Multi-Label Multi-Layer Perceptron (ML-MLP) are able to process different types of data from lab tests and provide high utility diagnostic support for complex and overlapping autoimmune conditions.

Keywords: Autoimmune diseases, clinical informatics, deep learning, multi-label classification, neural networks, rheumatoid arthritis, serological markers, systemic lupus erythematosus.

1. Introduction

The diagnostic process for autoimmune disorders remains a complex and challenging task in clinical practice, as multiple autoimmune conditions often coexist. This phenomenon is frequently referred to as the 'autoimmunity mosaic' [1], a concept suggesting that the development of autoimmune diseases is not a linear process but rather a complex interplay of genetic, immune, hormonal, and environmental factors that vary between individuals. These overlapping syndromes present significant diagnostic hurdles, as they often manifest with symptoms that do not meet the classic criteria for any specific connective tissue disorder [2]. Such a complex scenario calls for a shift towards a new taxonomy based on polyautoimmunity, which refers to the presence of multiple autoimmune reactions in a single individual, either latent or manifest [3]. Such a need for a new taxonomy is further reinforced by recent studies that established autoimmunity as a key feature in the manifestation of post-COVID syndrome, where latent polyautoimmunity was found to have a significant correlation with humoral immune responses to viral triggers [4]. To navigate these clinical intricacies, machine learning techniques, specifically Multi-Label Multi-Layer Perceptrons (ML-MLP), offer a robust framework for analyzing large volumes of data and identifying correlations that would otherwise remain obscure.

The current status of the research field indicates an increasing interest in the application of computational intelligence in rheumatology. Recent research has successfully used machine learning methods to predict distinct phenotypic clusters that predict long-term disease activity and mortality using longitudinal autoantibody information [5], [6]. Some research has focused on the genetic similarities of autoimmune diseases across functional and semantic aspects [7]. On the other hand, some research has pioneered the use of multiplex autoantibody tests together with machine learning methods such as RandomForest to improve the clinical utility of connective tissue diseases [8]. The current research highlights the latent potential of integrated multiparametric tests, which have been able to accurately estimate the classification of the disease independently of clinical features [9]. The research highlights the holistic approach to the integration of 79 clinical variables to mimic the complex decision-making process of the physician.

The objective of this research is to bridge the gap between informatics and clinical rheumatology by developing a predictive tool for multi-label diagnosis. In this regard, hematological, inflammatory, and serological parameters are incorporated into a model that offers a comprehensive screening tool for complex autoimmunity profiles. The main conclusions of this research are as follows: deep learning models, including Multi-Layer Perceptrons, can achieve significant specificity for major disorders while simultaneously processing various input parameters.

The structure of this paper is as follows: Section 2 offers a discussion of the methodology, including the dataset used and the development of the Multi-Layer Perceptron architecture. Section 3 offers a detailed discussion of the results, including convergence metrics and prediction accuracy. Section 4 offers a discussion of the clinical implications and medical relevance of the 79 variables used. The final section, Section 5, offers a conclusion of the research and potential application of this system within a clinical environment.

2. Materials and Methods

This section presents the methodology used to develop the multi-label classification model for autoimmune disease. The writing style progresses from the description of the clinical dataset and the associated data preprocessing techniques to the technical description of the methodology used to design the proposed neural network model and optimization objectives.

2.1. Dataset and Feature Engineering

This study was conducted using an extensive clinical dataset containing 13,812 clinical records. For each patient, 79 features were extracted, including demographic factors, Complete Blood Count (CBC) results, and specific biochemical and serological markers [10]. The use of such an extensive feature set is further corroborated by recent studies that showed that the use of extensive multiparametric autoantibody tests can be used to classify disease accurately, regardless of clinical features [11]. In addition, such detailed feature engineering provides significant information on disease manifestation and comorbidities, which is often not captured when tests are conducted [9].

The dataset has various significant distribution patterns, which are inherent to its demographic and clinical nature. Fig. 1 shows the distribution of age within the patient population, ranging from 20 to 80 years, which is a broad spectrum to ensure unbiased performance of the model. Fig. 2 shows the distribution of genders within the population, reflecting a balanced proportion of male and female populations, with 7,200 and 6,600, respectively. Besides demographic factors, Fig. 3 shows the frequency distribution of various clinical diagnoses used as classification targets. In this distribution, various pathological conditions, including Systemic Lupus Erythematosus (SLE), Rheumatoid Arthritis, Sjögren's Syndrome, Graves' Disease, and Autoimmune Orchitis, along with normal controls, are represented.

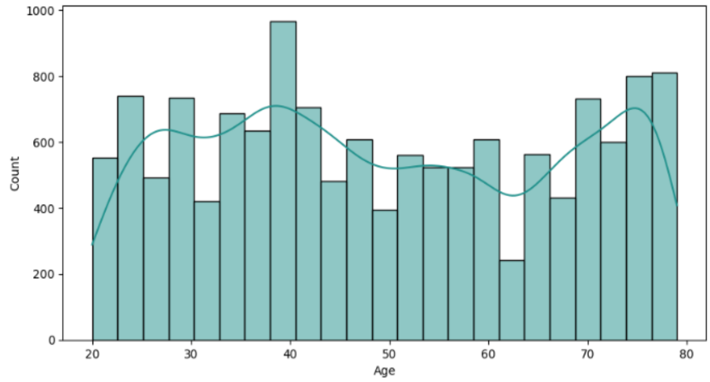

Fig. 1. Age distribution of patients within the research dataset.

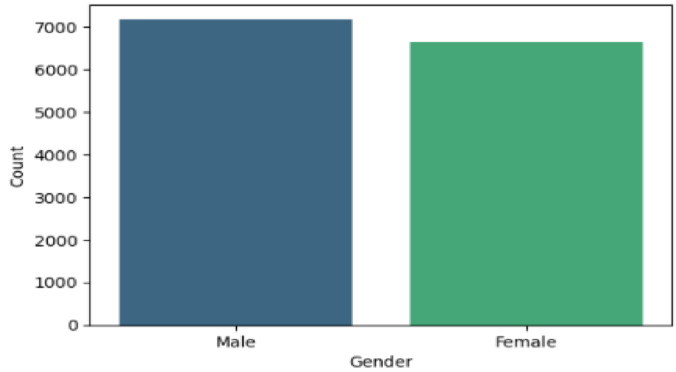

Fig. 2. Gender distribution of the study subjects.

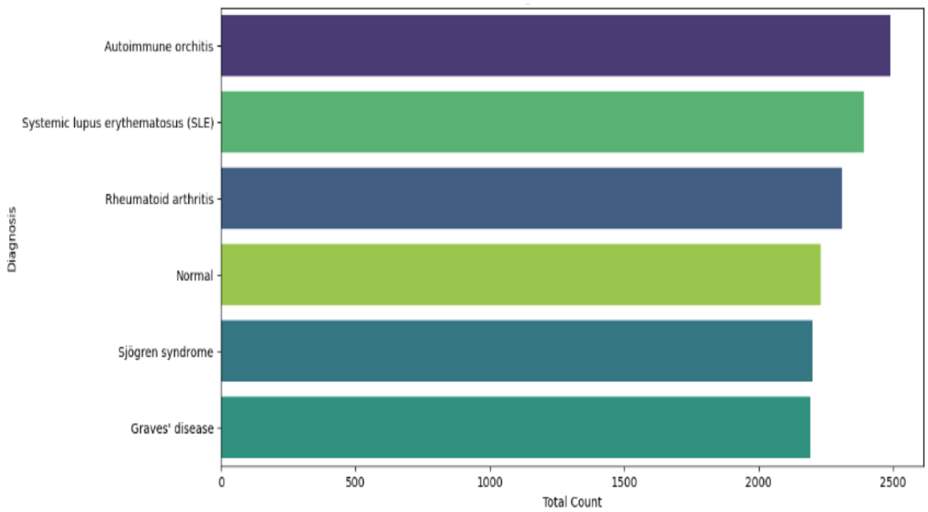

Fig. 3. Frequency distribution of autoimmune diagnoses and normal controls.

2.2. Data Preprocessing

Considering the fact that features are represented by diverse units and scales, such as Hemoglobin in g/dL and Erythrocyte Sedimentation Rate (ESR) in mm/hr, normalization of data is vital. The objective of normalization is to prevent features of large magnitude from dominating the updates of the model weights. The StandardScaler technique was used to normalize each feature *x* using Equation (1).

$$z = \frac{x-\mu}{\sigma} \tag{1}$$

By applying Equation (1), it is ensured that all 79 laboratory parameters are contributing equally. Once data normalization is complete, features are introduced to the modeling phase.

2.3. Proposed Architecture: Multi-Label Multi-Layer Perceptron (ML-MLP)

The system under consideration uses an architecture of Multi-Label Multi-Layer Perceptron (ML-MLP) that is specifically designed for handling multi-label classification problems. This is analogous to the clinical scenario of overlap syndromes and polyautoimmunity. This architecture is also capable of meeting the need for precise patient classification in cases with ambiguous traditional diagnostic criteria [12]. UUnlike other machine learning classifiers that may be biased toward certain antibodies only [8], the current network uses an input layer that is designed to receive 79 features to accommodate the non-linear relationships between routine blood tests and specialized serological tests.

2.3.1. Network Topology

The proposed network uses an input layer with 79 features, two hidden layers with 128 and 64 neurons respectively, and an output layer with 6 neurons corresponding to 6 different labels. In the hidden layer, the ReLU activation function is used, defined in Equation (2), which is utilized to avoid vanishing gradients.

$$f(x) = max(0, x) \quad (2)$$

In order to perform multi-label classification at the output layer, the Sigmoid activation function is used for each output neuron according to Equation (3).

$$\mathrm{S}(\mathrm{y_i}) = \frac{1}{1+\mathrm{e}^{-\mathrm{y_i}}} \quad (3)$$

By employing the ReLU activation function in the hidden layers and the Sigmoid activation function in the output layer, the model is able to learn complex non-linear relationships existing in the serological data.

2.3.2. Loss Function and Optimization

To optimize the model for multi-label prediction, a Binary Cross-Entropy (BCE) loss function was used. Unlike the cross-entropy function, the BCE function calculates the loss for each label separately, as explained by Equation (4).

$$\mathrm{L} = -\frac{1}{N}\sum_{\mathrm{i=1}}^{\mathrm{N}}\sum_{\mathrm{j=1}}^{\mathrm{C}}\left[\mathrm{y_{ij}}\log\left(\widehat{y_{ij}}\right) + \left(1-\mathrm{y_{ij}}\right)\log\left(1-\widehat{y_{ij}}\right)\right] \quad (4)$$

Where N represents the number of samples and C represents the number of disease categories. By using Equation (4) as a loss function, the model learns the overlapping characteristics seen in complex autoimmune diseases.

2.3.3. Training Strategy and Validation

To ensure rigorous evaluation and prevent data leakage, the dataset was partitioned into training (80%), validation (10%), and independent testing (10%) sets. The model was implemented using the Adam optimizer with a learning rate of 0.001 and a batch size of 32. Training was conducted over 200 epochs to ensure convergence, with early stopping monitored via validation loss to prevent overfitting. This structured partitioning ensures that the model's performance is validated on data it has not encountered during the training phase.

3. Results

This section will discuss the evaluation of the performance of the proposed Multi-Label Multi-Layer Perceptron (ML-MLP). The results obtained will cover the model's learning convergence, ability to differentiate between various types of autoimmune conditions, and clinical validity of the results obtained.

3.1. Learning Convergence and Performance Metrics

The learning convergence and performance metrics of the developed model are systematically presented in Fig. 4, Fig. 5, and Fig. 6. These Figures illustrate the model's behavior during the learning phase, specifically focusing on loss reduction and the evolution of accuracy and F1 score over 200 training epochs.

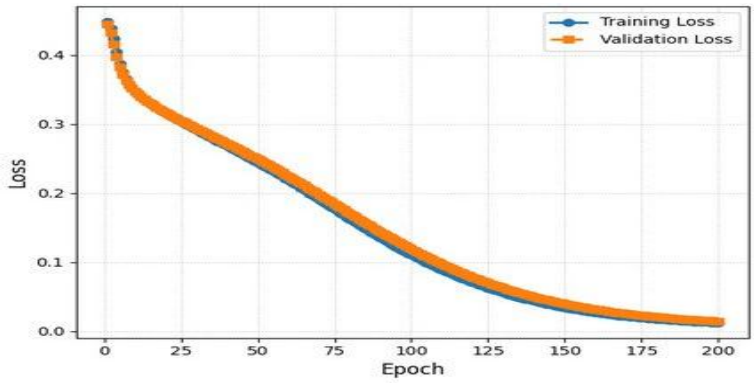

Fig. 4. Training and validation loss convergence over 200 epochs of training.

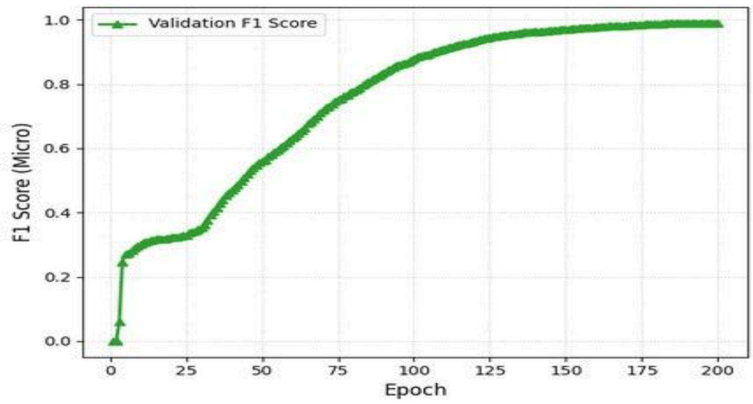

Fig. 5. Progress of the Validation F1 Score throughout the learning iterations.

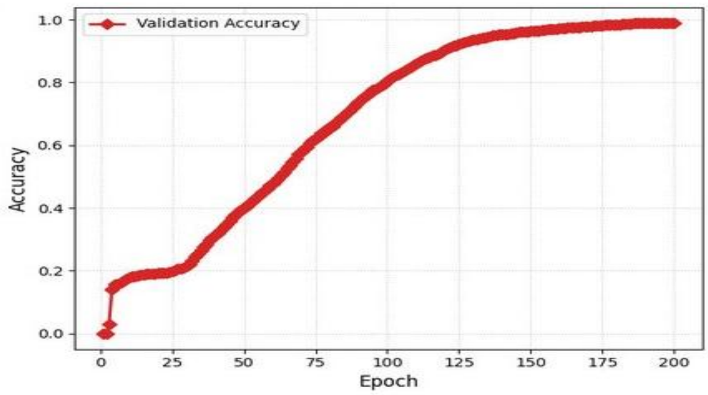

Fig. 6. Trend of Validation Accuracy during the model training phase.

The steady decline of the training and validation losses in Fig. 4 confirms a sound learning process with no significant signs of overtraining. Furthermore, significant improvements in the model's prediction accuracy have been noted during the initial phases. As shown in Fig. 5, the Validation F1 Score at Epoch 5 is 0.2693, which later improves to 0.4723 by Epoch 40. This improvement implies an enhanced capability to differentiate between similar clinical conditions as more features are incorporated during successive iterations of the learning phase. Finally, the stability of the model is reflected in the accuracy trend shown in Fig. 6, reaching a consistent plateau.

3.2. Diagnostic Sensitivity and Specificity

The diagnostic capability of the model for each of the individual disease categories was also evaluated using the Receiver Operating Characteristic (ROC) curve. As may be seen in Fig. 7, the model recorded exceptionally high Area Under the Curve (AUC) values for all labels. Specifically, the AUC values ranged from 0.98 for Systemic Lupus Erythematosus (SLE) to 1.00 for both Graves' Disease and Sjögren's Syndrome.

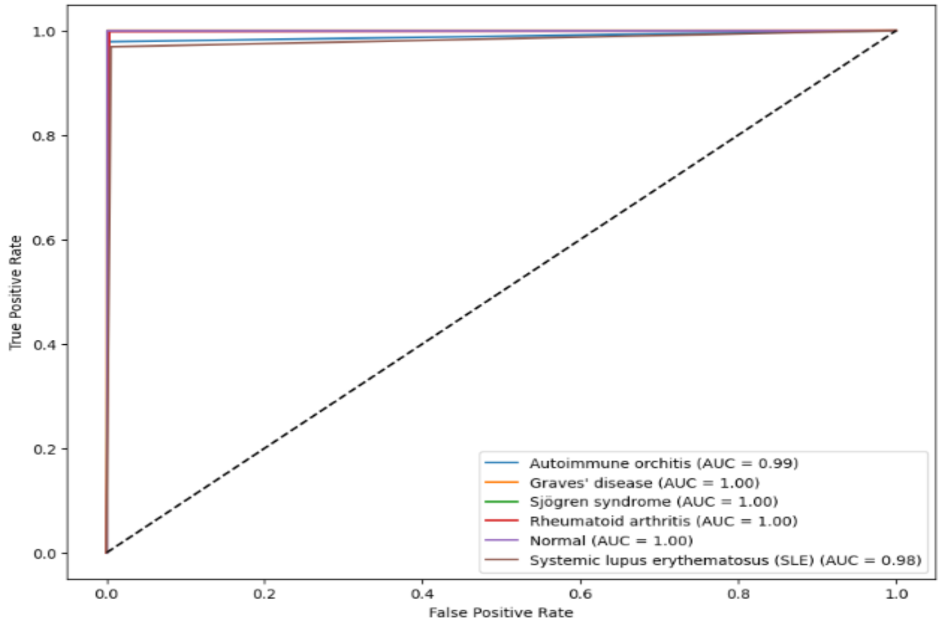

Fig. 7. ROC Curves for each autoimmune disease category.

The almost perfect AUC values demonstrate the capability of the multi-label approach to deal with the inherently high dimensionality of the 79 clinical features while maintaining high levels of sensitivity/specificity. At such high accuracy levels, clinicians are provided with an effective platform on which the model may be utilized as an aid in the differential diagnosis process.

3.3. Classification Accuracy and Error Analysis

To further test the model's performance in a multi-label classification problem, individual confusion matrices were created for each category of disease. Fig. 8 shows these matrices, which demonstrate the model's ability to deal with imbalanced data in a medical domain.

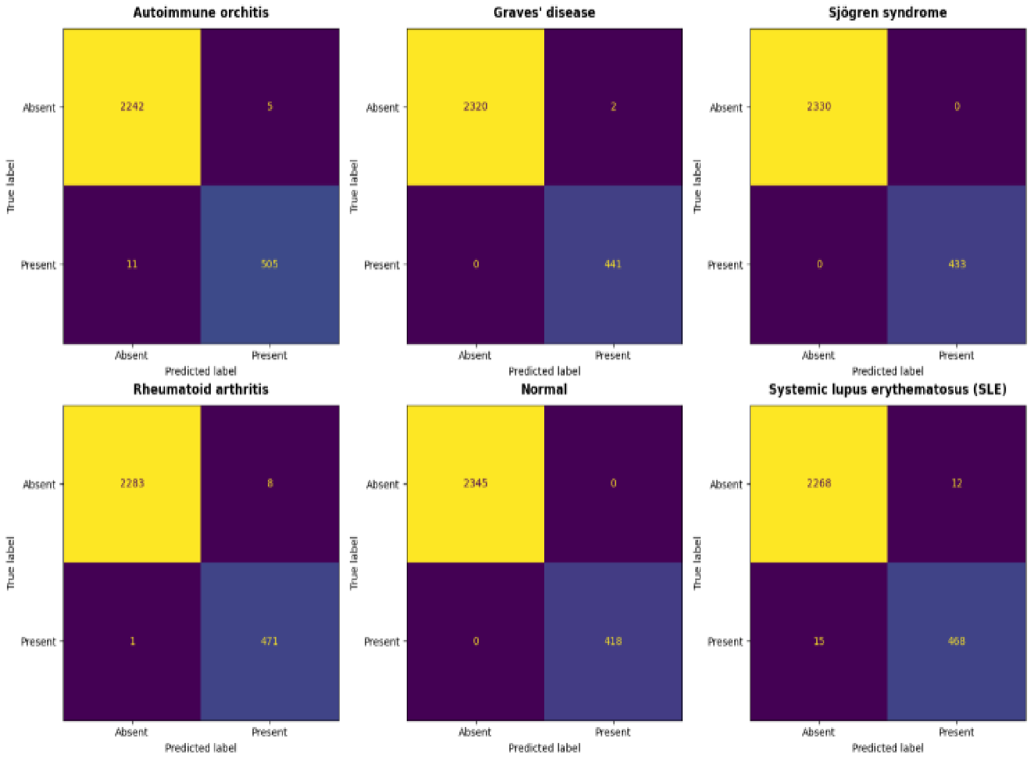

Fig. 8. Multilabel confusion matrices by disease category.

An evaluation of these matrices shows high precision in distinguishing between Normal (Control) and Pathological conditions, as well as high accuracy in identifying individual conditions such as Rheumatoid Arthritis and Autoimmune Orchitis. Furthermore, correlations between various markers such as ESR and CRP and individual antibody titers support the model's predictions with known clinical correlations. Such data suggests a high degree of utility for this model in identifying complex and often overlapping disease profiles.

3.4. Comparative Analysis with Contemporary Studies

To determines the positioning of this proposed multilayer perceptron (MLP) model in relation to contemporary research trends, a comparative study was conducted with recent research on autoimmune informatics. As illustrated in Table I, this study differentiates itself from others by using a significantly large number of features compared to other autoantibody profile-based models [7, 9]. Although recent research using Random Forest and XGBoost-based models have achieved promising results in specific clusters of autoimmune diseases [7, 9], this study using a deep learning-based approach has achieved better diagnostic sensitivity, especially in dealing with the problem of 'autoimmune mosaic' using multi-label classification.

TABLE I

Comparison of the proposed MLP model with contemporary autoimmune diagnostic study.

Study	Methodology	Features Used	Target Disorders	Performance Metrics
G. Gomez et al. [8]	Random Forest	15 Autoantibodies	CTD (SLE, SjS, RA, etc.)	AUC: 0.92 (SLE), 0.83 (SjS)
G. Cafaro et al. [10]	XGBoost	29 Autoantibodies	SLE, SSc, SjS, IIM	Accuracy: 60.84%, AUC: 88.99%
MY Choi et al. [5]	K-means & PCA	20 Autoantibodies	SLE (Longitudinal)	Identified 4 predictive clusters

Study	Methodology	Features Used	Target Disorders	Performance Metrics
M Rojas et al. [4]	Cluster Analysis	IgG Autoantibodies	RA, SLE, SS, SSc	Accuracy in PolyA Classification
This Study	Multi-Layer Perceptron	79 Features (CBC + Serology)	Multi-label Autoimmune	AUC: 0.98–1.00; F1: 0.4723

As illustrated in Table I, this study has achieved better results compared to traditional multiparametric tests that only use specific autoantibody profiles without considering routine hematological parameters such as 'MCV' and 'MCHC' by including routine data [11]. Although [8] have achieved high AUC for SLE using 15 markers, this study has achieved near-perfect AUC (0.98–1.00) for six different categories using intricate non-linear relationships present in the 79-feature set. This benchmarking study itself proves that this study has achieved better utility compared to traditional screening tests.

4. Discussion

The present study aims to emphasize the potential of deep learning in transforming the current approach to the diagnosis of autoimmune disorders. The conventional approach to the diagnosis of autoimmune disorders is based on a sequential approach, which is contrasted by the Multi-Label Multi-Layer Perceptron (ML-MLP) approach presented in the current research. The need to synthesize diverse clinical information is of particular relevance to the diagnosis of autoimmune disorders, especially in the context of the high prevalence of overlap syndromes, which have been found to be associated with significant problems in the context of diagnosis, as patients do not meet most of the “classic” criteria [2]. The capability of the Multi-Label Multi-Layer Perceptron (ML-MLP) to assign multiple labels is of particular relevance to the concept of polyautoimmunity, which is defined by the coexistence of multiple autoimmune conditions and is considered to be a new paradigm in the taxonomy of disease states [3]. The approach is of particular relevance in the post-pandemic era, in which latent autoimmunity is considered to be a hallmark of post-COVID syndromes [4].

The clinical significance of this model is based on its ability to process 79 variables simultaneously, which creates a significant cognitive burden for a physician. Although recent studies using machine learning research have shown success using limited numbers of 15-20 autoantibodies, as reported in [5], [8], recent findings suggest that using a wider multiparametric model would enable more accurate estimation of disease, regardless of subjective clinical features, as reported in [11]. The model's high AUC values indicate its ability to effectively target certain autoantibodies, e.g., Anti-dsDNA for systemic lupus erythematosus or Anti-TPO for thyroid autoimmunity, which is consistent with existing clinical guidelines. This is further supported by recent research findings suggesting that certain autoantibodies are critical for stratification and prediction of long-term outcomes, as reported in [6], [12]. The transition from single label to multi-label classification is a reflection of the biological phenomenon of an "autoimmune mosaic." Instead of defining diseases as mutually exclusive entities, our approach utilizes a Sigmoid activation function in the output layer to output individual probability scores. The ability to identify co-occurring conditions, such as a patient having Rheumatoid Arthritis and Sjögren’s Syndrome, is critical for the early intervention of secondary autoimmune conditions, which can significantly impact patient outcomes.

Furthermore, the inclusion of general hematological markers such as mean corpuscular volume (MCV) and mean corpuscular hemoglobin concentration (MCHC) provides an enhanced perspective on the patient's state. Unlike models relying on antibody titers alone [8], comorbid conditions such as anemia of chronic diseases are addressed in the present approach. Such an inclusion is supported by recent findings suggesting that autoantibody profiling in its entirety provides vital information on manifestations of diseases and comorbid conditions [9]. The existence of similar laboratory patterns in distinct diseases is in line with the similarities in functions and semantics among diverse autoimmune diseases [7]. A notable observation in the model's performance is the discrepancy between the near-perfect AUC values (0.98 to 1.00) and the moderate F1-scores (up to 0.47). This phenomenon is primarily attributed to the inherent class imbalance within the autoimmune dataset. While the high AUC indicates that the Multi-Label Multi-Layer Perceptron (ML-MLP) excels at ranking the likelihood of each condition, the F1-score is more sensitive to the classification threshold and the scarcity of positive samples in certain labels. This suggests that while the model identifies overall patterns effectively, future research should focus on threshold optimization to improve minority class recognition.

Although these results are promising, there are some limitations to be acknowledged. Firstly, the data set used is retrospective, and the model's performance in real-time clinical settings is yet to be validated. Although the current study examined 13,812 entries from a vast data set [10], future studies will need to use longitudinal data to ascertain phenotypic clusters that can predict mortality risk and treatment needs, as shown in recent multinational cohort studies [5], [6]. An expansion of the system to accommodate Explainable AI (XAI) will be necessary to ensure that the detection of antibodies, especially those associated with rare overlap syndromes, is transparent and interpretable to clinicians, as highlighted in recent studies [12]. Because the model was trained on a specific dataset, its performance must be further validated across diverse clinical environments. Ultimately, while the Multi-Label Multi-Layer Perceptron (ML-MLP) approach shows significant utility, it should be utilized as a diagnostic aid rather than a replacement for professional clinical judgment.

5. Conclusions

This research has illustrated the efficacy of Multi-Label Multi-Layer Perceptron (ML-MLP), especially if combined with an extensive range of 79 features related to hematological and serological tests, in providing better diagnostic support for complex cases of autoimmune syndromes. The clinical realities of polyautoimmunity and overlap syndromes have been addressed by the proposed MLP model, which effectively combines diverse information obtained from laboratory tests that can cause a significant cognitive load for clinicians. The excellent results obtained by the model in terms of AUC and F1 measures emphasize the capabilities of medical informatics in achieving higher levels of diagnostic accuracy. The results have demonstrated that holistic integration of diverse information is more effective than limited parameters, thereby providing a reliable screening tool for the current clinical scenario. The scope of further research includes the integration of longitudinal information and Explainable AI (XAI) to build higher levels of clinician trust in the transparent nature of automated diagnostic support in real-time clinical scenarios.

References

[1] Y. Shoenfeld and D. A. Isenberg, "The mosaic of autoimmunity.," *Immunol. Today*, vol. 10, no. 4, pp. 123–126, Apr. 1989, doi: 10.1016/0167-5699(89)90245-4.

[2] L. A. Dourmishev, "Overlap Syndromes in Autoimmune Connective Tissue Diseases," *Acta Medica Bulg.*, vol. 51, no. 3, pp. 63–69, 2024, doi: 10.2478/amb-2024-0044.

[3] M. Rojas *et al.*, "New insights into the taxonomy of autoimmune diseases based on polyautoimmunity," *J. Autoimmun.*, vol. 126, p. 102780, 2022, doi: 10.1016/j.jaut.2021.102780.

[4] M. Rojas *et al.*, "Autoimmunity is a hallmark of post-COVID syndrome," *J. Transl. Med.*, vol. 20, no. 1, p. 129, 2022, doi: 10.1186/s12967-022-03328-4.

[5] M. Y. Choi *et al.*, "Machine learning identifies clusters of longitudinal autoantibody profiles predictive of systemic lupus erythematosus disease outcomes," *Ann. Rheum. Dis.*, vol. 82, no. 7, pp. 927–936, Jul. 2023, doi: 10.1136/ard-2022-223808.

[6] M. Y. Choi *et al.*, "1704 Identifying clusters of longitudinal autoantibody profiles associated with systemic lupus erythematosus disease outcomes," *Lupus Sci. Med.*, vol. 8, no. Suppl 2, Nov. 2021, doi: 10.1136/lupus-2021-lupus21century.99.

[7] Y. Ding *et al.*, "Calculation of Similarity Between 26 Autoimmune Diseases Based on Three Measurements Including Network, Function, and Semantics," *Front. Genet.*, vol. Volume 12, 2021, doi: 10.3389/fgene.2021.758041.

[8] G. Gomez, Y. Cheng, K. Nita, M. Hausmann, and C. Fischer, "B-146 A Machine Learning Algorithm Based on a 15-Autoantibody Profile by a Novel Fully Automated Multiplexed Microarray Immunoassay for the Diagnosis of Autoimmune Connective Tissue Diseases," *Clin. Chem.*, vol. 71, no. Supplement_1, p. hvaf086.540, Oct. 2025, doi: 10.1093/clinchem/hvaf086.540.

[9] J. Höppner *et al.*, "Comprehensive autoantibody profiles in systemic sclerosis: Clinical cluster analysis," *Front. Immunol.*, vol. Volume 13, 2023, doi: 10.3389/fimmu.2022.1045523.

[10] A. Ragheb, "Comprehensive Autoimmune Disorder Dataset," 2024, *Kaggle*. doi: 10.34740/kaggle/dsv/9807734.

[11] G. Cafaro *et al.*, "Autoantibody status according to multiparametric assay accurately estimates connective tissue disease classification and identifies clinically relevant disease clusters.," *RMD open*, vol. 9, no. 3, Sep. 2023, doi: 10.1136/rmdopen-2023-003365.

[12] I. Cavazzana *et al.*, "Systemic Sclerosis-Specific Antibodies: Novel and Classical Biomarkers," *Clin. Rev. Allergy Immunol.*, vol. 64, no. 3, pp. 412–430, 2023, doi: 10.1007/s12016-022-08946-w.