American Journal of BioScience
2023; 11(6): 142-149
http://www.sciencepublishinggroup.com/j/ajbio
doi: 10.11648/j.ajbio.20231106.12
ISSN: 2330-0159 (Print); ISSN: 2330-0167 (Online)

Construction of a Clinical Prediction Model for Systemic Sclerosis Cuproptosis-Related Genes Using Machine Learning

Xinmin Huang†, Xu Cai†, Xinpeng Chen, Yiwei Hong, Zhengbo Yan, Jianwei Xiao*

Rheumatology, Shenzhen Futian Hospital for Rheumatic Diseases, Shenzhen, China

Email address:

108812365@qq.com (Jianwei Xiao)

*Corresponding author

† Xinmin Huang and Xu Cai are co-first authors.

To cite this article:

Xinmin Huang, Xu Cai, Xinpeng Chen, Yiwei Hong, Zhengbo Yan et al. (2023). Construction of a Clinical Prediction Model for Systemic Sclerosis Cuproptosis-Related Genes Using Machine Learning. *American Journal of BioScience*, *11*(6), 142-149.
https://doi.org/10.11648/j.ajbio.20231106.12

Received: October 31, 2023; **Accepted**: November 15, 2023; **Published**: November 24, 2023

Abstract: *Objective:* This study aims to identify cuproptosis-related genes in Systemic Sclerosis (SSc) and construct a clinical prediction model. *Methods:* The GSE33463 dataset was retrieved from the GEO database, and gene set enrichment analysis (GSEA) was used to analyze the expression of pathways related to cuproptosis. Cuproptosis-related genes were extracted, and potential key genes for SSc were selected using the LASSO and Boruta methods to construct a clinical prediction model. The model's predictive ability was evaluated using K-nearest neighbors (KNN) and Lightgbm methods, with assessment based on ROC curves, PR curves, confusion matrices, F-values, and 5-fold cross-validation. The importance of model variables was evaluated using SHAP analysis. *Results:* Cuproptosis-related pathways were upregulated in SSc. Four key cuproptosis-related genes (PDHB, DLST, PDHA1, DBT) were identified using the LASSO and Boruta methods, leading to the construction of a clinical prediction model through multivariable logistic regression. The model exhibited a C-index of 0.91, an AUC of 0.914 under the ROC curve, and strong performance in 5-fold cross-validation. KNN and Lightgbm models achieved AUC values of 0.9243 and 0.9763, respectively. PR curve AUC values of 0.8492 and 0.9480 demonstrated high precision, while confusion matrix results revealed KNN and Lightgbm model accuracies of 0.8663 and 0.932, respectively. The models provide a basis for the early diagnosis of SSc. *Conclusion:* The clinical prediction model, based on four cuproptosis-related genes, demonstrates high predictive capability, aiding in the early diagnosis of SSc patients.

Keywords: Systemic Sclerosis, Cuproptosis, Machine Learning, Prediction Model, Confusion Matrix

1. Introduction

Systemic sclerosis (SSc), also known as scleroderma, is a rare connective tissue disease characterized by immune dysfunction, microvascular abnormalities, and fibrosis affecting the skin and internal organs [1, 2]. The global prevalence of SSc is estimated to be approximately 17.6 per 100,000 individuals, with an annual incidence rate of 1.4 per 100,000 people [3]. SSc can impact multiple systems, and research suggests that early diagnosis leads to a longer life expectancy, with a median survival period of 12 years for SSc patients [4]. Therefore, exploring molecular biomarkers related to the pathogenesis of SSc is beneficial for improving the early diagnosis rate of the disease.

Cuproptosis represents a novel non-apoptotic cell death pathway distinct from other known regulatory cell death mechanisms such as apoptosis, necroptosis, autophagy, and ferroptosis. It is a newly discovered, copper-dependent, and regulatable mode of cell death. Cuproptosis involves the targeted accumulation of copper ions within mitochondria, where they bind to lipoylated proteins in the tricarboxylic acid cycle of mitochondrial respiration. This binding drives the aggregation of lipoylated-modified proteins, subsequently leading to the downregulation of iron-sulfur cluster proteins, inducing protein toxicity stress, and ultimately resulting in cell death [5]. The functional role of cuproptosis in SSc remains

elusive. Therefore, there is an urgent need for a deeper understanding of SSc and the development of novel biomarkers and treatment modalities.

This study employs bioinformatics analysis to identify crucial cuproptosis-related genes, establish a clinical predictive model for diagnosing SSc, and validate the model using various machine learning methods. The findings from this research may significantly contribute to the early detection of SSc.

2. Materials and Methods

2.1. Data Retrieval

Using "systemic sclerosis" as the keyword, the GSE33463 dataset was obtained from the Gene Expression Omnibus database, comprising 19 healthy controls and 41 SSc patients, with the detection platform being GPL16947.

2.2. Gene Set Enrichment Analysis (GSEA) and Acquisition of Cuproptosis-Related Gene Expression Profiles

Cuproptosis-related genes (NFE2L2, NLRP3, ATP7B, ATP7A, SLC31A1, FDX1, LIAS, LIPT1, LIPT2, DLD, DLAT, PDHA1, PDHB, MTF1, GLS, CDKN2A, DBT, GCSH, DLST) were identified. The GSEA software was utilized for the analysis of the cuproptosis signaling pathway, with the number of random permutations set at 2,000, and the statistical significance determined by a false discovery rate (FDR) < 0.05. The R software's limma package was used to extract the expression profiles of the 19 cuproptosis-related genes.

2.3. Identification of HUB Genes

The LASSO and Boruta methods were employed to screen for potential key genes associated with SSc pathogenesis. The intersection of the results obtained from both methods yielded the HUB genes associated with Cuproptosis in SSc.

2.4. Construction of Clinical Predictive Model

A predictive model was established using multivariable logistic regression, accompanied by the creation of a nomogram. The model's performance was assessed by measuring the C-index, plotting ROC curves and calibration plots, and depicting decision curve analysis to evaluate the net benefit at different threshold probabilities, thereby determining the clinical utility of the predictive model.

2.5. Machine Learning Model Validation

The K-Nearest Neighbors (KNN) and LightGBM algorithms were employed to validate the binary classification model built in step 1.4. The hyperparameters were optimized using a random grid set at 20. The model's predictive ability was evaluated using Receiver Operating Characteristic (ROC) curves, Precision-Recall (PR) curves, confusion matrices, F-values, and 5-fold cross-validation. The SHAP method was used to assess the importance of each variable in the model.

3. Results

3.1. GSEA Results

The results of Gene Set Enrichment Analysis (GSEA) revealed an upregulation of Cuproptosis-related pathways in SSc, with a false discovery rate (FDR) of less than 0.05 (Figure 1).

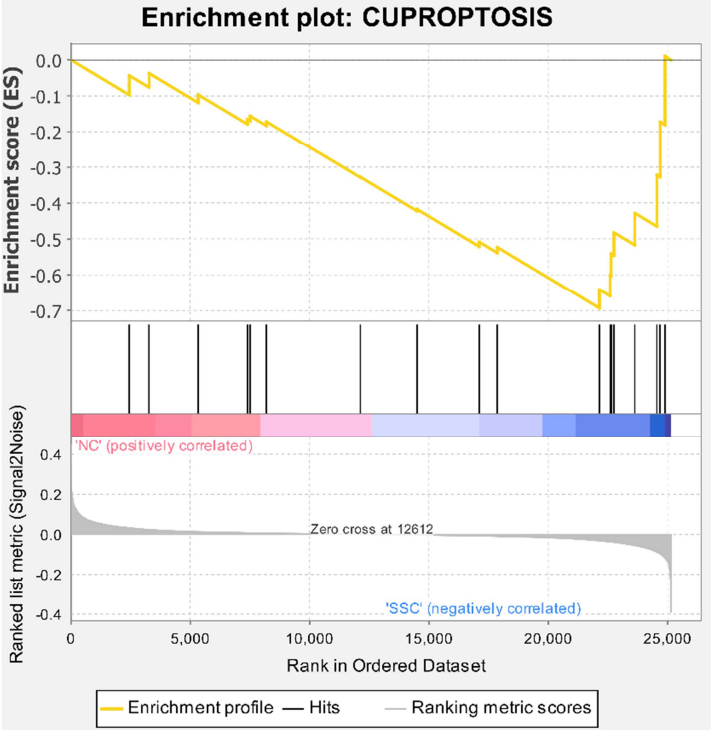

Figure 1. *GSEA Analysis Results.*

3.2. Identification of HUB Genes

Through the LASSO method, four key HUB genes (PDHB, PDHA1, LIPT1, FDX1) were selected, while the Boruta method identified five HUB genes (PDHB, DLST, GCSH, PDHA1, DBT) (Figure 2). Taking the intersection of these two sets of results yielded four common HUB genes (PDHB, DLST, PDHA1, DBT).

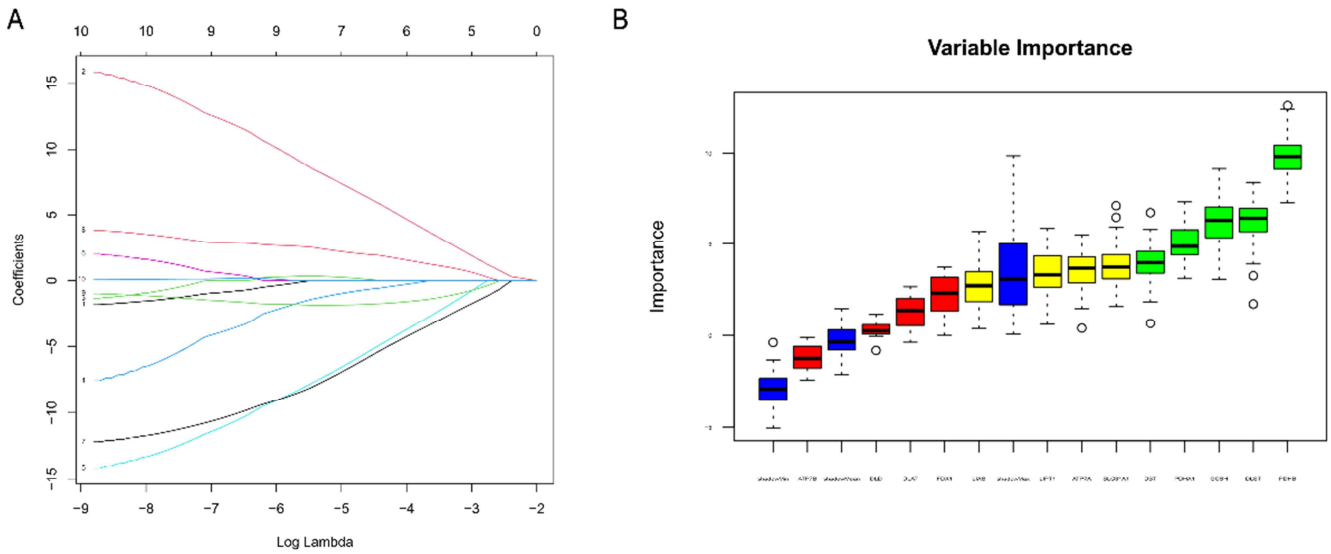

Figure 2. *Key Gene Selection (A. LASSO B. Boruta).*

3.3. Clinical Predictive Model Results

A clinical predictive model was constructed by incorporating these four HUB genes through multivariable logistic regression, visualized using a nomogram (Figure 3). The model achieved a C-index of 0.91. The ROC curve indicated an area under the curve (AUC) of 0.9114 (Figure 43A). The calibration curve displayed good consistency (Figure 4B). Calibration curve results demonstrated the highest clinical net benefit when the threshold probability ranged from 0.2 to 0.8 (Figure 5).

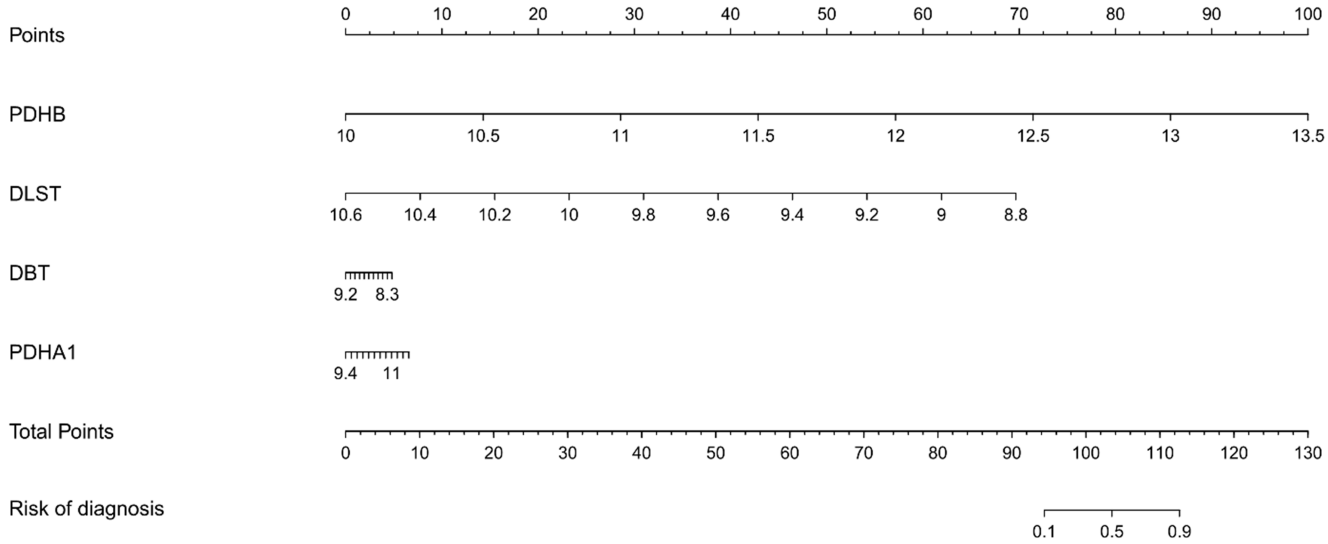

Figure 3. *Clinical Prediction Model Nomogram.*

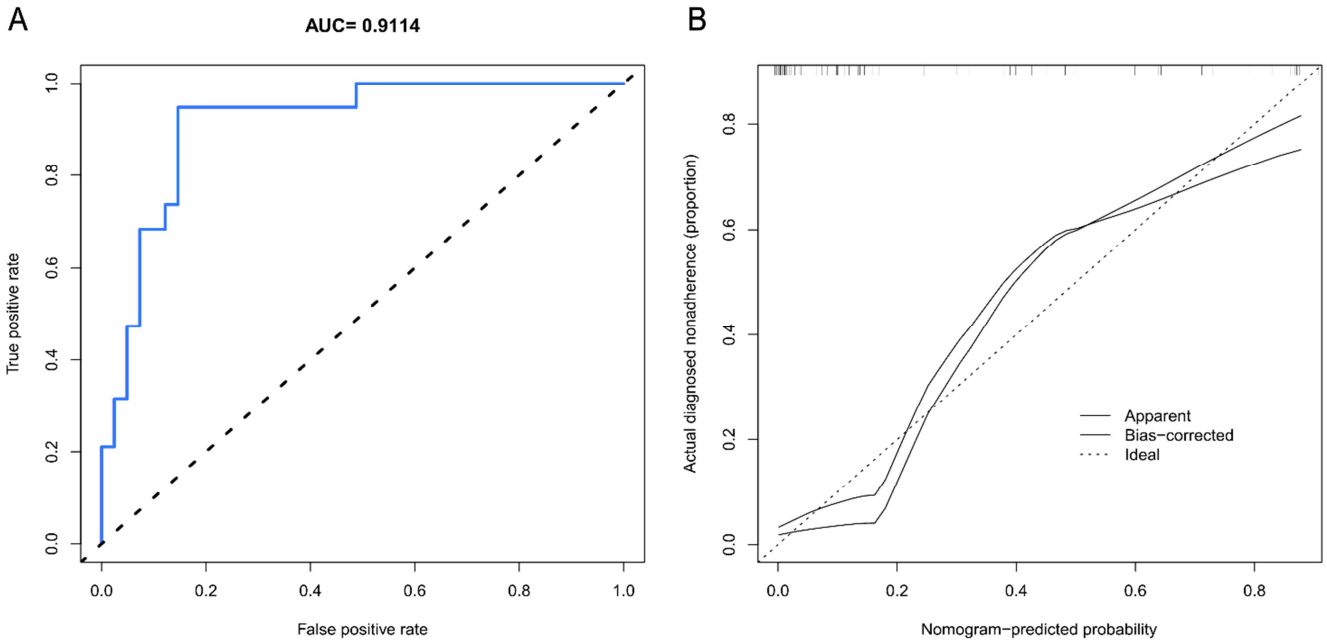

Figure 4. *ROC Curve of the Predictive Model and Clinical Decision Curve (A: Dashed line represents the ideal model, solid line represents the actual performance of the model, the closer the solid line is to the dashed line, the stronger the model's predictive ability. B: ROC curve with an AUC of 0.9114).*

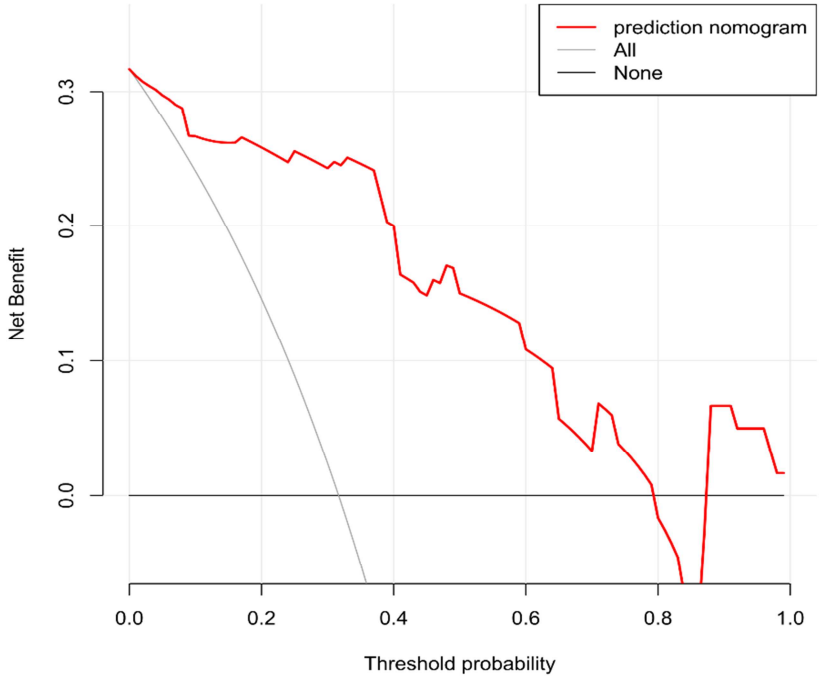

Figure 5. *Clinical Decision Curve.*

3.4. Machine Learning Validation Results

The results of the K-Nearest Neighbors (KNN) and LightGBM algorithms showed ROC areas under the curve of 0.9243 and 0.9763, respectively (Figure 6). The Precision-Recall (PR) curve areas under the curve were 0.8492 and 0.9480, respectively (Figure 7). Shap visualization revealed that in both KNN and LightGBM algorithms, the most important variables were DLST and PDHB (Figure 8), consistent with the clinical predictive model results. The confusion matrix displayed the following outcomes: For the KNN predictive model, True Positive Rate (TPR) was 0.739, True Negative Rate (TNR) was 0.946, accuracy was 0.8663, and the F1 score was 0.810. The LightGBM predictive model showed a TPR of 0.857, TNR of 0.974, accuracy of 0.932, and an F1 score of 0.899 (Figure 9). In 5-fold cross-validation, all model ROC values exceeded 0.7 (Figure 10).

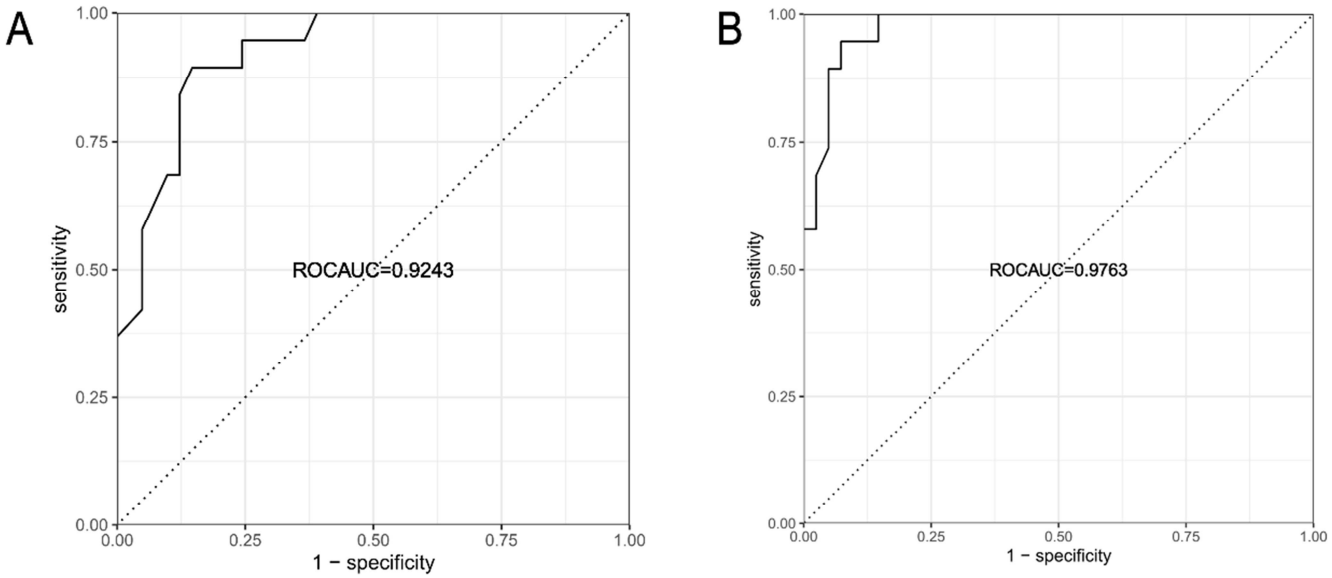

Figure 6. *ROC Curve (A. KNN B. Lightgbm).*

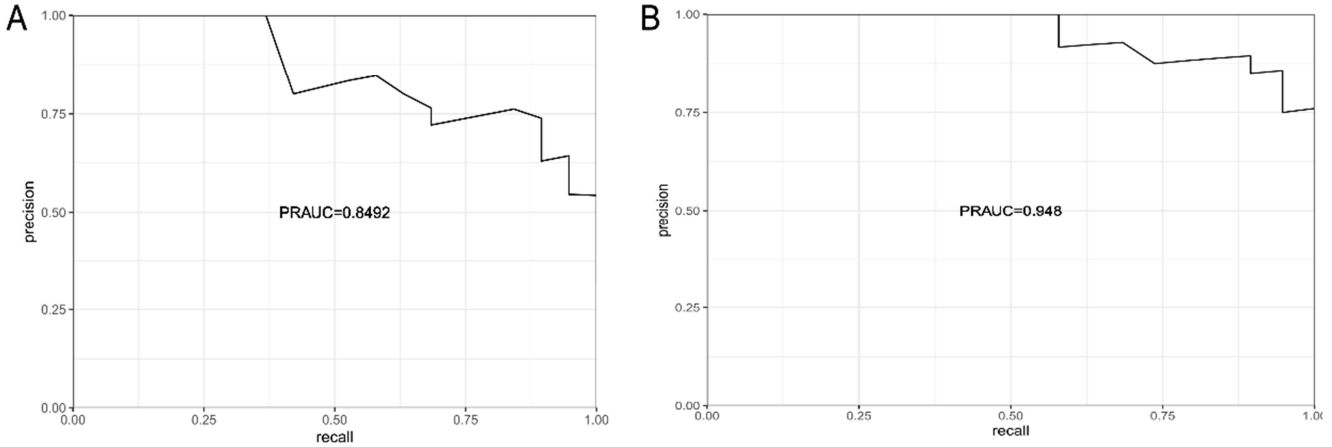

Figure 7. *PR Curve (A. KNN B. Lightgbm).*

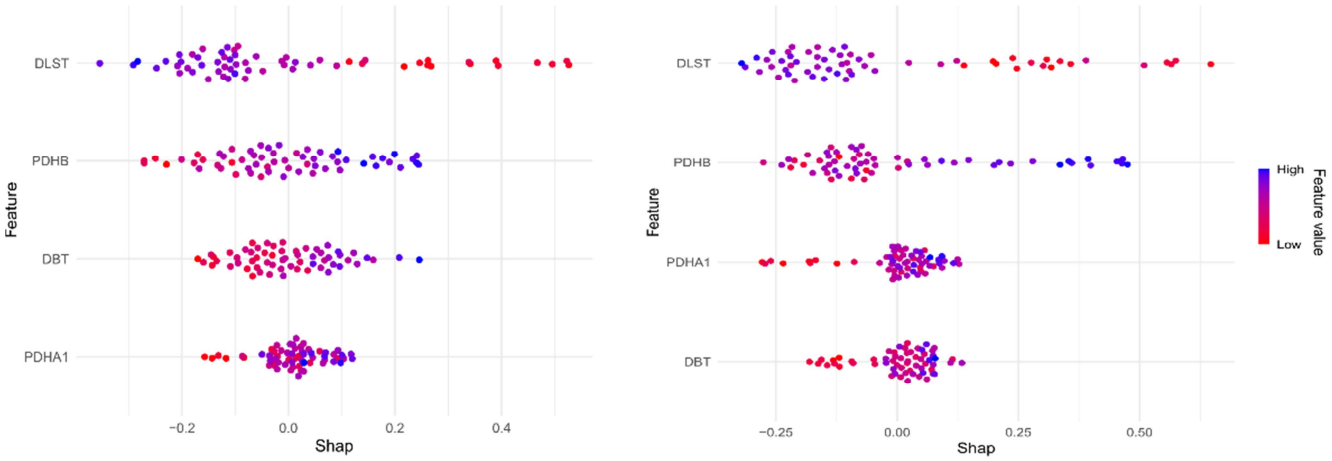

Figure 8. *SHAP Visualization (A. KNN B. Lightgbm).*

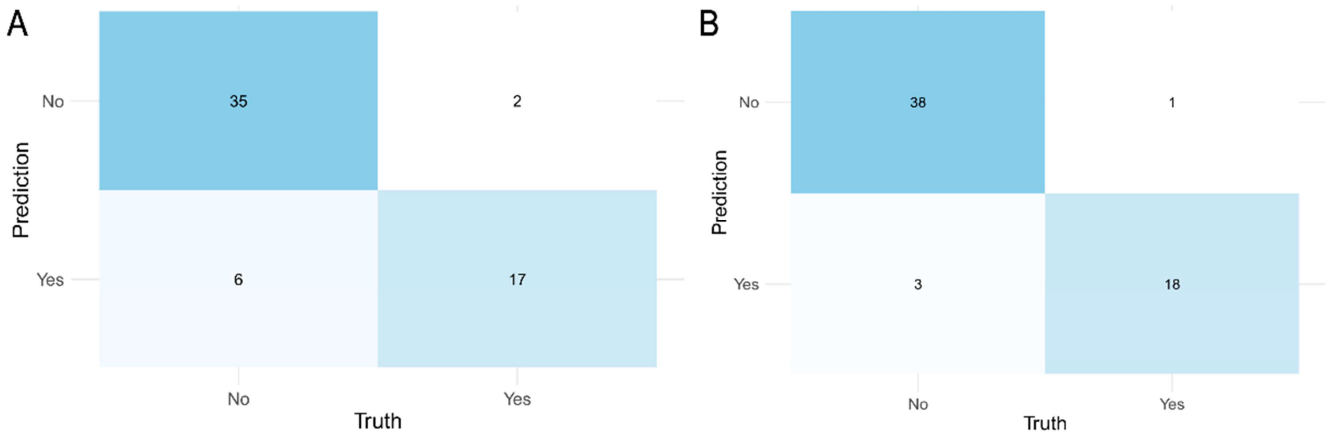

Figure 9. *Confusion Matrix (A. KNN B. Lightgbm).*

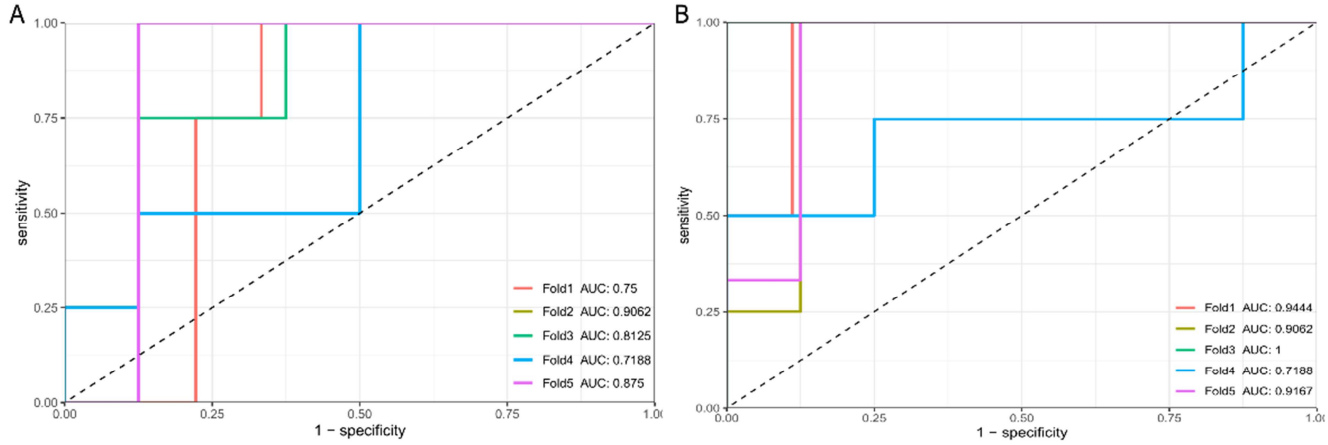

Figure 10. *Five-Fold Cross-Validation (A. KNN B. Lightgbm).*

4. Discussion

Systemic sclerosis (SSc) is a challenging autoimmune disease that significantly impacts patients, characterized by high morbidity and mortality rates. Investigating the pathogenic mechanisms of SSc and achieving early diagnosis for intervention are of paramount importance. Despite numerous etiological studies conducted on SSc, the immunological pathogenesis and relevant therapeutic targets remain elusive. Previous research has highlighted the significant role of cuproptosis in rheumatic immune diseases [6, 7]. However, the role of cuproptosis in the development and progression of SSc has not been previously studied.

Copper ions are crucial trace elements in the human body, playing pivotal roles in various biological processes, including energy metabolism, antioxidant defense, and cell signal transduction. Nevertheless, when copper ion concentrations exceed the normal range, oxidative stress can occur. This is because high concentrations of copper ions can catalyze the generation of reactive oxygen species (ROS), such as superoxide anions, hydrogen peroxide, and hydroxyl radicals. These ROS can damage biological macromolecules such as cell membranes, proteins, and DNA, leading to cellular injury and death [8]. Excessive ROS-induced oxidative stress plays a crucial role in the pathogenesis of systemic sclerosis. Firstly, oxidative stress can promote the activation and proliferation of fibroblasts, leading to increased collagen production and deposition, resulting in fibrosis of the skin, lungs, and other organs [9]. Secondly, oxidative stress can cause damage and dysfunction of vascular endothelial cells, leading to vasoconstriction and thrombosis, further exacerbating the symptoms of systemic sclerosis [10]. Moreover, oxidative stress can also influence the function of the immune system, causing abnormal activation of immune cells and the persistence of inflammatory reactions, thereby accelerating the progression of systemic sclerosis [11]. Studies have shown that copper ion levels are significantly elevated in SSc patients compared to healthy individuals [12]. Cuproptosis is characterized by the intracellular accumulation of copper ions, leading to the accumulation of mitochondrial lipoylated proteins and instability of Fe-S cluster proteins, ultimately resulting in cell death. Certain drugs such as penicillamine, a copper ion chelator, have been used to treat diseases associated with connective tissue, such as rheumatoid arthritis and scleroderma. This study revealed the upregulation of cuproptosis-related pathways in SSc patients, suggesting a potential role for cuproptosis in the pathogenesis of SSc.

In this study, we constructed a clinical predictive model for diagnosing SSc, with four cuproptosis-related genes identified as potential key genes involved in the pathogenesis of SSc. Machine learning results demonstrated that various indicators,

including ROC curves, calibration curves, PR curves, and the C-index, showed that the model effectively distinguishes between normal patients and SSc patients. Among the four cuproptosis-related genes, PDHB and PDHA1 represent two E1 subtypes of pyruvate dehydrogenase, primarily located in the mitochondria of cells, where they catalyze the conversion of glucose-derived pyruvate to acetyl coenzyme A [13, 14]. Acetyl coenzyme A is an indispensable component of the tricarboxylic acid (TCA) cycle [15], where it combines with oxaloacetate to form citrate, which subsequently enters the TCA cycle [16, 17]. Citrate is a critical molecule in energy production metabolism, and one study has shown a significant reduction in citrate levels in the serum of SSc patients compared to the control group [18]. Once synthesized within the mitochondria, citrate promotes oxidative phosphorylation, leading to ATP production [19]. Under inflammatory conditions, such as in systemic lupus erythematosus, the reduced availability of energy increases the demand for ATP [20]. Citrate has also been shown to possess important immunological properties, regulating proinflammatory processes within macrophages [18], suggesting that modulating citrate production through the regulation of PDHB and PDHA1 may represent a potential therapeutic approach for treating SSc. DLST encodes a crucial protein involved in mitochondrial lipoylation [5]. Studies by Udumula et al. have indicated that inhibiting DLST reduces oxidative phosphorylation in myeloid cells and modulates immune microenvironments [21]. Colliou et al. discovered through related research that increased expression of the DLAT gene significantly enhances intestinal Th17 cells, thus preventing inflammatory diseases [22]. DBT regulates the degradation of branched-chain amino acids and is a transacylase component of the mitochondrial multienzyme branched-chain α-ketoacid dehydrogenase complex. A study by Ahn et al. suggested that under hypoxic conditions, peroxiredoxin V binding to DBT can modulate cellular oxidative stress and mitochondrial transport, among other biological activities [23]. In summary, based on previous research, it is conjectured that the four cuproptosis-related genes may play a role in the pathogenesis of SSc by influencing oxidative stress, inflammation, and other pathways. However, further in-depth research is required to elucidate how they may prevent and treat SSc by regulating cuproptosis.

5. Conclusion

This study employed bioinformatics techniques to identify four core cuproptosis-related genes and construct a clinical predictive model for SSc, providing a theoretical basis for early SSc diagnosis. Nonetheless, elucidating the mechanisms of cellular cuproptosis in SSc remains a focus of future research, necessitating further cell and animal experiments for validation.

Funding

The present study was funded by the Chinese Medicine Research Project of Traditional Chinese Medicine Bureau Of Guangdong Province (20221342), and the Shenzhen Futian District Health and Public Welfare Research Project (grant no. FTWS2021063, FTWS2022021).

Author Contributions

Conceptualisation: Jianwei XIAO. Data curation: Xinmin Huang. Data analysis: Xinmin Huang, Xinpeng CHEN. Writing-original draft: Xinmin Huang, Xu Cai, HONG Yiwei, YAN Zhengbo. Approval of final manuscript: all authors.

ORCID

Xinmin Huang: https://orcid.org/0009-0003-3183-8869
Xu Cai: https://orcid.org/0000-0002-8553-1585
Xinpeng Chen: https://orcid.org/0009-0008-7455-3173
Yiwei Hong: https://orcid.org/0009-0006-6399-0277
Zhengbo Yan: https://orcid.org/0000-0002-7954-5358
Jianwei Xiao: https://orcid.org/0000-0002-1178-538X

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

We sincerely thank the contributors of the GEO database for making their data publicly available, which has facilitated this research.

References

[1] Korman B. Evolving insights into the cellular and molecular pathogenesis of fibrosis in systemic sclerosis. Transl Res. 2019; 209: 77-89. doi: 10.1016/j.trsl.2019.02.010.

[2] Denton CP, Khanna D. Systemic sclerosis. Lancet. 2017; 390 (10103): 1685-1699. doi: 10.1016/S0140-6736(17)30933-9.

[3] Bairkdar M, Rossides M, Westerlind H, et al M. Incidence and prevalence of systemic sclerosis globally: a comprehensive systematic review and meta-analysis. Rheumatology (Oxford). 2021; 60 (7): 3121-3133. doi: 10.1093/rheumatology/keab190.

[4] Pokeerbux MR, Giovannelli J, Dauchet L, et al. Survival and prognosis factors in systemic sclerosis: data of a French multicenter cohort, systematic review, and meta-analysis of the literature. Arthritis Res Ther. 2019; 21 (1): 86. doi: 10.1186/s13075-019-1867-16.

[5] Tsvetkov P, Coy S, Petrova B, et al. Copper induces cell death by targeting lipoylated TCA cycle proteins [published correction appears in Science. 2022 Apr 22; 376 (6591): eabq4855]. Science. 2022; 375 (6586): 1254-1261. doi: 10.1126/science.abf0529.

[6] Jiang M, Liu K, Lu S, et al. Verification of cuproptosis-related diagnostic model associated with immune infiltration in rheumatoid arthritis. Front Endocrinol (Lausanne). 2023; 14: 1204926. doi: 10.3389/fendo.2023.1204926.

[7] Li W, Guan X, Wang Y, et al. Cuproptosis-related gene identification and immune infiltration analysis in systemic lupus erythematosus. Front Immunol. 2023; 14: 1157196. doi: 10.3389/fimmu.2023.1157196.

[8] Svegliati S, Spadoni T, Moroncini G, Gabrielli A. NADPH oxidase, oxidative stress and fibrosis in systemic sclerosis. Free Radic Biol Med. 2018; 125: 90-97. doi: 10.1016/j.freeradbiomed.2018.04.554.

[9] Ames PRJ, Bucci T, Merashli M, et al. Oxidative/nitrative stress in the pathogenesis of systemic sclerosis: are antioxidants beneficial?. Free Radic Res. 2018; 52 (10): 1063-1082. doi: 10.1080/10715762.2018.1525712.

[10] Doridot L, Jeljeli M, Chêne C, Batteux F. Implication of oxidative stress in the pathogenesis of systemic sclerosis via inflammation, autoimmunity and fibrosis. Redox Biol. 2019; 25: 101122. doi: 10.1016/j.redox.2019.101122.

[11] Abdulle AE, Diercks GFH, Feelisch M, et al. The Role of Oxidative Stress in the Development of Systemic Sclerosis Related Vasculopathy. Front Physiol. 2018; 9: 1177. Published 2018 Aug 24. doi: 10.3389/fphys.2018.01177.

[12] Li X, Wang X, Li J, et al. Serum Concentrations of Trace Elements/Minerals in Patients with Diffuse Systemic Sclerosis. Biol Trace Elem Res. 2021; 199 (7): 2440-2443. doi: 10.1007/s12011-020-02368-5.

[13] Zhu Y, Wu G, Yan W, et al. miR-146b-5p regulates cell growth, invasion, and metabolism by targeting PDHB in colorectal cancer. Am J Cancer Res. 2017; 7 (5): 1136-1150. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5446479/

[14] Saunier E, Benelli C, Bortoli S. The pyruvate dehydrogenase complex in cancer: An old metabolic gatekeeper regulated by new pathways and pharmacological agents. Int J Cancer. 2016 Feb 15; 138 (4): 809-17. doi: 10.1002/ijc.29564.

[15] Echeverri Ruiz NP, Mohan V, Wu J, et al. Dynamic regulation of mitochondrial pyruvate metabolism is necessary for orthotopic pancreatic tumor growth. Cancer Metab. 2021 Nov 8; 9 (1): 39. doi: 10.1186/s40170-021-00275-4.

[16] Feng X, Zhang L, Xu S, et al. ATP-citrate lyase (ACLY) in lipid metabolism and atherosclerosis: An updated review. Prog Lipid Res. 2020 Jan; 77: 101006. doi: 10.1016/j.plipres.2019.101006.

[17] Patel MS, Nemeria NS, Furey W, et al. The pyruvate dehydrogenase complexes: structure-based function and regulation. J Biol Chem. 2014 Jun 13; 289 (24): 16615-23. doi: 10.1074/jbc.R114.563148.

[18] Infantino V, Iacobazzi V, Menga A, et al. A key role of the mitochondrial citrate carrier (SLC25A1) in TNFα- and IFNγ-triggered inflammation. Biochim Biophys Acta. 2014 Nov; 1839 (11): 1217-1225. doi: 10.1016/j.bbagrm.2014.07.013.

[19] Iacobazzi V, Infantino V. Citrate--new functions for an old metabolite. Biol Chem. 2014 Apr; 395 (4): 387-99. doi: 10.1515/hsz-2013-0271.

[20] Ouyang X, Dai Y, Wen JL, et al. ^{1}H NMR-based metabolomic study of metabolic profiling for systemic lupus erythematosus. Lupus. 2011 Nov; 20 (13): 1411-20. doi: 10.1177/0961203311418707.

[21] Udumula MP, Sakr S, Dar S, et al. Ovarian cancer modulates the immunosuppressive function of CD11b+Gr1+ myeloid cells via glutamine metabolism. Mol Metab. 2021 Nov; 53: 101272. doi: 10.1016/j.molmet.2021.101272.

[22] Colliou N, Ge Y, Sahay B, et al. Commensal Propionibacterium strain UF1 mitigates intestinal inflammation via Th17 cell regulation. J Clin Invest. 2017 Nov 1; 127 (11): 3970-3986. doi: 10.1172/JCI95376..

[23] Ahn SH, Yang HY, Tran GB, et al. Interaction of peroxiredoxin V with dihydrolipoamide branched chain transacylase E2 (DBT) in mouse kidney under hypoxia. Proteome Sci. 2015 Feb 5; 13: 4. doi: 10.1186/s12953-014-0061-2.