



OPEN ACCESS

EDITED BY

Zhibin Lv,
Sichuan University, China

REVIEWED BY

Alex Jung,
Aalto University, Finland
Jiawei Wang,
Yale University, United States

*CORRESPONDENCE

Haitao Yang,
✉ 17702488036@163.com
Zhenyuan Wang,
✉ wzy1322@aliyun.com

RECEIVED 23 May 2023

ACCEPTED 02 August 2023

PUBLISHED 31 August 2023

CITATION

Dong H, Du Z, Ma H, Zhou Z, Yang H and Wang Z (2023), Prediction of distinct populations of innate lymphoid cells by transcriptional profiles.
Front. Genet. 14:1227452.
doi: 10.3389/fgene.2023.1227452

COPYRIGHT

© 2023 Dong, Du, Ma, Zhou, Yang and Wang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Prediction of distinct populations of innate lymphoid cells by transcriptional profiles

Haiyao Dong^{1,2}, Zhenguang Du³, Haoming Ma⁴, Zhicheng Zhou³, Haitao Yang^{2*} and Zhenyuan Wang^{1,2*}

¹Department of Thoracic Surgery, China Medical University, Shenyang, China, ²Department of Thoracic Surgery, The People's Hospital of Liaoning Province, Shenyang, China, ³Department of No. 3 Oncology, The People's Hospital of Liaoning Province, Shenyang, China, ⁴College of Software, Northeastern University, Shenyang, China

Innate lymphoid cells (ILCs) are a unique type of lymphocyte that differ from adaptive lymphocytes in that they lack antigen receptors, which primarily reside in tissues and are closely associated with fibers. Despite their plasticity and heterogeneity, identifying ILCs in peripheral blood can be difficult due to their small numbers. Accurately and rapidly identifying ILCs is critical for studying homeostasis and inflammation. To address this challenge, we collect single-cell RNA-seq data from 647 patients, including 26,087 transcripts. Background screening, Lasso analysis, and principal component analysis (PCA) are used to select features. Finally, we employ a deep neural network to classify lymphocytes. Our method achieved the highest accuracy compared to other approaches. Furthermore, we identified four genes that play a vital role in lymphocyte development. Adding these gene transcripts into model, we were able to increase the model's AUC. In summary, our study demonstrates the effectiveness of using single-cell transcriptomic analysis combined with machine learning techniques to accurately identify congenital lymphoid cells and advance our understanding of their development and function in the body.

KEYWORDS

innate lymphoid cells, machine learning, DNN, LASSO, gene expression

Highlights

- Our study demonstrates the feasibility of combining machine learning methods with feature extraction models for cell immunotyping.
- To compare various classification models and feature extraction methods, we conducted comparative experiments and determined that the optimal model was the combination of DNN and LASSO.
- Our findings indicate that the incorporation of the four genetic information found in the literature can enhance the accuracy of the classification model.

1 Introduction

In the past decade, innate lymphoid cells (ILCs) have garnered significant attention from researchers due to their crucial role in the innate immune system (Hazenbergh and Spits, 2014; Artis and Spits, 2015; Eberl et al., 2015; Vivier et al., 2018). These heterogeneous lymphocytes originate from lymphoid progenitor cells in the bone marrow. Notably, they lack the

rearrangement of antigen-specific receptors that depend on recombination activation genes, do not express antigen-specific receptors unique to acquired immune cells, and do not exhibit surface markers similar to those found on other immune cells. The transcription factors and secreted cytokines necessary for the development of various ILCs differ. Based on these factors and cytokines, ILCs can be classified into natural killer (NK) cells, ILC1s, ILC2s, ILC3s, and others (Montaldo et al., 2015; Miller et al., 2018; Vacca et al., 2019). ILCs are predominantly found in barrier regions such as the respiratory tract, digestive tract mucosa, and skin. They respond to local cytokine signals in their microenvironment and serve early immune surveillance and regulatory functions by secreting cytokines and other mediators. ILCs also act as a bridge between innate and acquired immunity, regulating systemic immune responses by coordinating the functions of acquired immune cells.

The accurate classification of ILCs is of great medical significance, as the functions and behaviors of various ILCs in the body can vary greatly. Abnormalities in the functions of ILCs can impact the onset and progression of various conditions, including inflammation, autoimmune diseases, metabolic disorders, and allergies (Elemam et al., 2017; Li et al., 2017; Golebski et al., 2019; Bartemes and Kita, 2021; Kabata et al., 2022; Kumar, 2022; Surace and Wilhelm, 2022). To effectively utilize the potential of ILCs in disease diagnosis and treatment, it is crucial to classify ILCs with the highest possible accuracy. Misclassification can lead to the use of inappropriate treatment strategies in clinical trials (Everaere et al., 2016; Everaere et al., 2018; Kogame et al., 2022). Despite their significant role, there is currently no effective method for identifying different cell subpopulations among ILCs. At present, the identification of ILCs primarily rely on flow cytometry, which uses specific antibodies to identify surface markers of different types of ILCs. However, as there are many types of ILCs, there often be some overlap in the functions of each subgroup. Therefore, the accuracy of marker-based identification techniques needs to be improved. In addition, although new sequencing technologies are developing (Hu et al., 2023), revealing patterns of gene expression at the cellular level, these new sequencing technologies are often costly. All in all, there is a need for accurate and low-cost methods for identifying ILCs.

In recent years, computational methods have been widely applied to mine biological information using omics data (Tyanova et al., 2016; Hériché et al., 2019; Efremova and Teichmann, 2020; Kaur et al., 2021; Badia-i Mompel et al., 2022; Watson et al., 2022). The expansion of genomic, proteomic, transcriptomic, and metabolomic data has provided an amazing opportunity for the application of machine learning methods. Researchers have developed a large number of tools, methods, and resources to fully utilize these data for precision medicine.

In this paper, we obtained genes associated with innate immune cells and used their expression levels to predict immune typing. Differential gene expression levels can directly reflect the developmental state of cells, and changes in gene expression can also affect the levels of proteins and metabolites. This work has yielded a precise and low-cost approach to classifying ILCs.

2 Methods

In this section, we provided a detailed overview of the implementation of this work, which includes the framework, data

preprocessing, feature compression, and evaluation. To construct our machine learning model, we needed three main components: data points, features, and labels (Jung, 2022). In this study, we built our model using the expression matrix of innate immune cells obtained from the GEO database, where data points represent individual cells, features represent the intensity of gene expression, and labels indicate the immune subtype that each cell corresponds to. With these components, we were able to develop and evaluate our machine learning models.

2.1 Workflow

Firstly, we obtained single-cell transcriptome data associated with innate lymphoid cells from the GEO database (Barrett et al., 2012), and then obtained known genes associated with innate lymphoid cells from DisGeNet (Piñero et al., 2020). We performed Related genes, PCA and LASSO analysis using the expression of genes associated with innate lymphoid cells as features and extracted the most significant features related to immune typing. In addition, in their latest study, Korchagina et al. identified *STATS*, *BATF*, *IKAROS*, *RUNX3*, *C-MAF*, *BCL11B*, and *ZBTB46* as genes closely associated with Innate Lymphoid Cells. And *STATS*, *IKAROS*, and *C-MAF* have been included already in our previous gene set. Therefore, we incorporated the expression of four additional hub genes in our subsequent analysis. After feature dimensionality reduction, we inputted all these features into a deep neural network (DNN). The DNN displayed immune typing based on important gene expression features. The workflow of our method is shown in Figure 1.

2.2 Datasets processing

For our study, we utilized the Björklund dataset (Björklund et al., 2016), which comprises sequencing data of lymphoid cells from three independent human volunteers. We conducted a comprehensive search of ILCs-related datasets across multiple databases. Unfortunately, many of these datasets were incomplete, with inadequate or absent classification of ILCs. However, the Björklund dataset, which is from human tonsil, emerged as a standout resource. Published in *Nature Immunology*, it has been widely cited and is considered authoritative. Therefore, we chose it to perform downstream analysis. In this dataset, the specimens for transcriptional analysis came from three donors: a 56-year-old (Donor A), a 44-year-old (Donor B), and a 23-year-old (Donor C). All procedures were carried out at the Karolinska University Hospital, located in Huddinge. The regional ethical committee at the Karolinska Institutet granted approval for the collection of these anonymous tissue samples. Informed consent was provided by all patients, or by their legal guardians in cases where the patients were under 18 years of age. We selected 647 samples annotated with immune typing, including 74 NK cells, 126 ILC1 cells, 139 ILC2 cells, and 308 ILC3 cells. Each sample in this dataset contains 26,087 transcripts. We mapped these transcripts to their corresponding genes using the GRCh37 assembly of the human

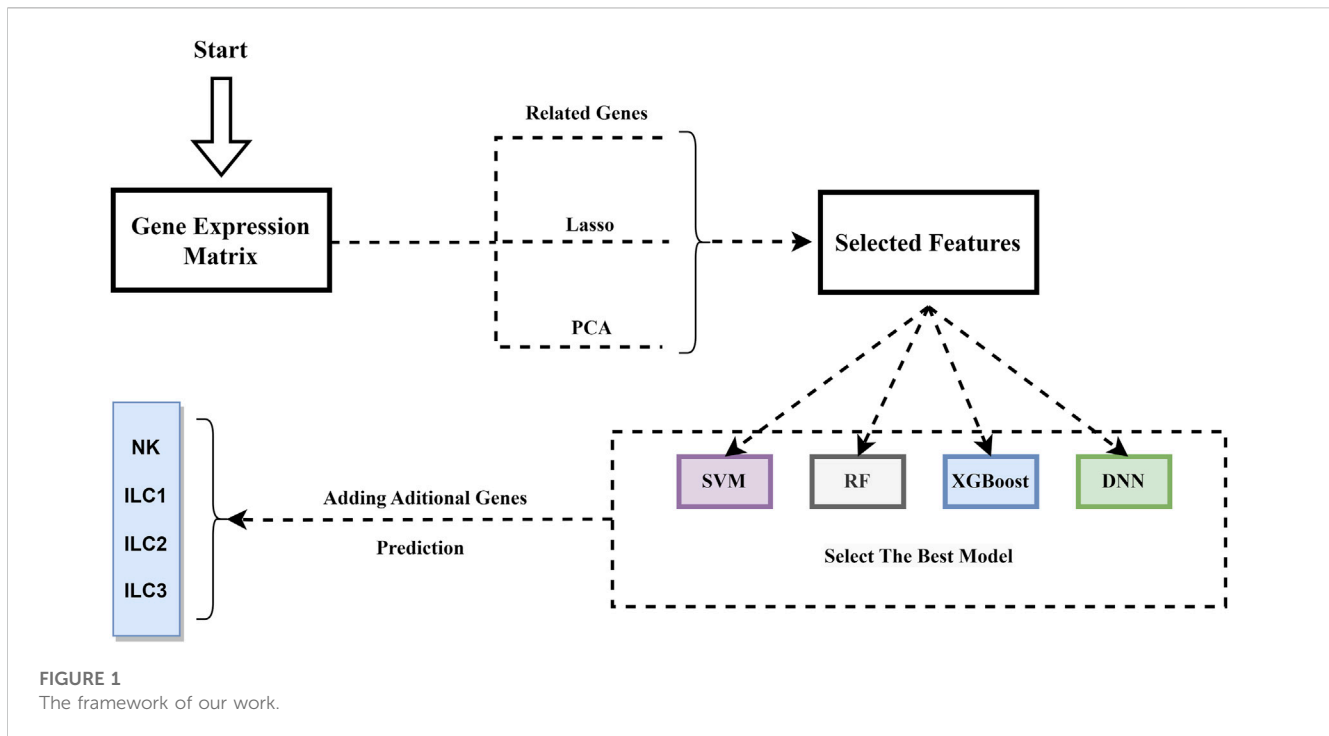


FIGURE 1 The framework of our work.

genome. Additionally, we obtained 292 genes associated with innate lymphoid cells from DisGeNet. Moreover, our analysis of additional studies revealed that BCL11B, BATF, RUNX3, and ZBTB46 are associated with lymphoid cell immune typing (Korchagina et al., 2023).

2.3 Feature compression

This section presents a comparative analysis of three commonly used feature compression methods. For each method, we provide detailed information about its implementation.

- Related Genes

To reduce the dimensionality of features using the related gene method, we followed a specific process. First, we obtained a gene set related to the disease from the DisGeNet website. Then, we intersected this gene set with the gene set of the original data to obtain a new, reduced gene set.

- PCA

Principal component analysis (PCA) is a commonly used data dimensionality reduction technique, which aims to transform high-dimensional data into lower-dimensional data for better data processing and analysis (Daffertshofer et al., 2004). The basic idea of PCA is to map the original data to a new coordinate system via linear transformation, such that the mapped data has the maximum variance. This linear transformation is achieved by computing the covariance matrix of the data and its eigenvectors. Specifically, assume that we have an $n \times p$ data matrix X , where n is the number of samples and p is the number of variables. We first

need to center X by subtracting the mean of each variable from the entire variable, resulting in a new matrix \tilde{X} , where,

$$\tilde{X}_{ij} = X_{ij} - \frac{1}{n} \sum_{k=1}^n X_{kj} \tag{1}$$

Next, we compute the covariance matrix S of \tilde{X} , i.e.,

$$S = \frac{1}{n-1} \tilde{X}^T \tilde{X} \tag{2}$$

Then, we perform eigendecomposition on S to obtain the eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ and their corresponding eigenvectors v_1, v_2, \dots, v_p . Finally, we select the top k eigenvectors v_1, v_2, \dots, v_k and project the data X onto these eigenvectors to obtain a new $n \times k$ matrix Y , where $Y_{ij} = v_j^T X_i$. In this way, we complete PCA dimensionality reduction, transforming the original $n \times p$ data matrix X into a new $n \times k$ matrix Y , where k is the number of variables we choose.

- LASSO

Lasso (Least Absolute Shrinkage and Selection Operator) is a widely used linear regression technique for feature selection and sparse modeling (Osborne et al., 2000). Lasso constrains the model parameters using L1 regularization, which shrinks some parameters to zero, achieving feature selection. Specifically, we can represent Lasso regression using the following formula:

$$\min_{\beta} \frac{1}{2n} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 \tag{3}$$

Here, β is the model parameter vector, X is the feature matrix, y is the response variable vector, and λ is the regularization parameter. The L1 regularization term $\|\beta\|_1$ shrinks some parameters to zero, achieving feature selection and model sparsity. By adjusting the regularization parameter λ , we can control the degree of sparsity of the model.

2.4 DNN four-class classification

A Four-class deep neural network (DNN) can be defined as a function $f: \mathbb{R}^n \rightarrow \mathbb{R}^4$ that maps an input vector $\mathbf{x} \in \mathbb{R}^n$ to an output vector $\mathbf{y} \in \mathbb{R}^4$. This function can be represented as a composition of layers, where each layer i consists of a linear transformation $\mathbf{z}^{(i)} = \mathbf{W}^{(i)}\mathbf{h}^{(i-1)} + \mathbf{b}^{(i)}$ and a nonlinear activation function $g^{(i)}$, with $\mathbf{h}^{(0)} = \mathbf{x}$ as the input and $\mathbf{h}^{(i)} = g^{(i)}(\mathbf{z}^{(i)})$ as the output of layer i . The weights and biases of each layer are denoted by $\mathbf{W}^{(i)}$ and $\mathbf{b}^{(i)}$, respectively. The process can be formulated as:

$$\mathbf{z}^{(L)} = \mathbf{W}^{(L)}\mathbf{h}^{(L-1)} + \mathbf{b}^{(L)} \quad (4)$$

In the last layer, we apply the softmax function, as below:

$$\mathbf{y} = \text{softmax}(\mathbf{z}^{(L)}) \quad (5)$$

To transform the output $\mathbf{z}^{(L)}$ into a probability vector \mathbf{y} , where L is the index of the last layer. During training, we use the cross-entropy loss function, as:

$$\mathcal{L}(\mathbf{y}, \mathbf{t}) = - \sum_{i=1}^4 t_i \log y_i \quad (6)$$

To measure the difference between the predicted probability vector \mathbf{y} and the true label vector \mathbf{t} , where $t_i \in \{0, 1\}$ indicates whether the i -th class is the correct class.

The training process we use backpropagation to compute gradients and update weights and biases. At each training iteration, we feed the input vector \mathbf{x} to the DNN, compute the output vector \mathbf{y} , compare it with the true label vector \mathbf{t} , then use backpropagation to compute the gradients and use optimization algorithms such as gradient descent to update weights and biases.

2.5 Evaluating metrics

Accuracy is a performance metric used to evaluate classification models, which represents the proportion of correctly classified samples to the total number of samples.

Specifically, given a classification model with predicted labels \hat{y}_i and corresponding true labels y_i , the accuracy of the model can be calculated using the following formula:

$$\text{Accuracy} = \frac{\sum_{i=1}^n \mathbb{1}(\hat{y}_i = y_i)}{n} \quad (7)$$

Here, n is the total number of samples, and $\mathbb{1}$ is the indicator function that takes the value 1 when $\hat{y}_i = y_i$ and 0 otherwise. The numerator of the formula represents the sum of correctly classified samples, and the denominator represents the total number of samples, i.e., the proportion of correctly classified samples to the total number of samples.

2.6 Parameter

To ensure the reproducibility of our experiments, we provide detailed information on the model parameters used in this work. We implemented PCA and Lasso dimensionality reduction methods using the sklearn library. To ensure fairness in dimensionality

reduction, all three methods were reduced to the same dimension 200. So PCA was set 200 principal components, while the alpha value in the Lasso model was set to 0.1. For the four classification models, three machine learning models were built using sklearn library, while the DNN was designed manually. The network layer parameters were set to (512, 64, 4), and the learning rate was set to 0.001.

3 Results

In this section, we conducted a thorough analysis of the sample size to ensure its appropriateness. Subsequently, we evaluated various methods for feature dimensionality reduction and machine learning classification, and selected the optimal model. Finally, we examined the impact of incorporating hub genes and validated their effectiveness.

3.1 Sample size verification

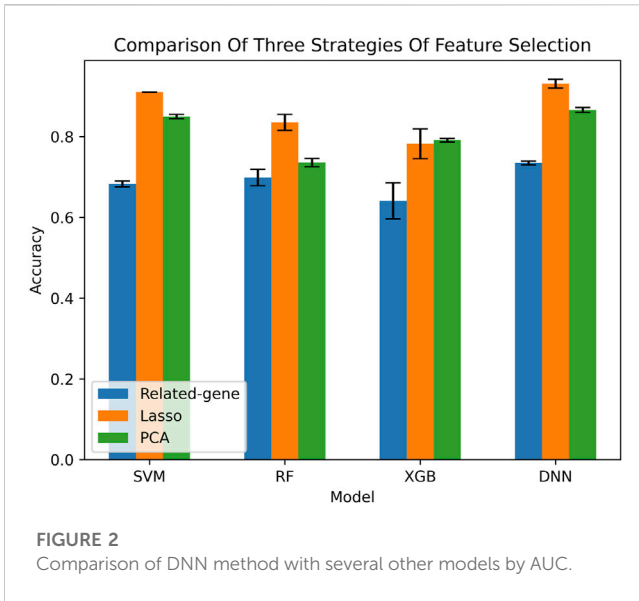
Our analysis involved raw data from the GEO database, which included properties of immune typing. To ensure clear labeling, we assigned the labels 0, 1, 2, and 3 to NK cells, ILC1, ILC2, and ILC3, respectively. This resulted in a dataset containing four types of samples, consisting of 74 NK cells, 126 ILC1 cells, 139 ILC2 cells, and 308 ILC3 cells. Importantly, the sample was relatively balanced among the different types of cells.

To assess the statistical performance of our data, we used an online web server called SSizer (Li et al., 2020). And the result illustrates that our data met Type 3 statistical indicators, indicating that our sample size was sufficient for our analysis. Furthermore, this demonstrates that when the sample size exceeded 300, the overlap was 0.5, indicating that our sample size was appropriate.

3.2 Comparative results

To provide a comprehensive evaluation of our method, we compared it with several other methods. Since there is no consensus on the best feature selection method, we simultaneously compared three different feature selection methods. The feature selection process is described in detail in Section 2 of our study. In addition, our method is based on deep neural networks (DNN) and compared with several other common machine learning classification models, including support vector machines, XGBoost, and random forests (Wang and Hu, 2005; Chen et al., 2015; Biau and Scornet, 2016). We tested each classification model using the three methods of feature selection, resulting in a total of 12 groups of classification results.

To assess the accuracy of our method, we performed 5-fold cross-validation. This is a common machine learning model evaluation method where the dataset is randomly divided into five mutually exclusive subsets. Each subset is used once as a validation set while the remaining four subsets are used for training. This process is repeated five times, with each subset used once as the validation set, to evaluate the performance of the model. The results, as shown in Figure 2, unequivocally demonstrate that the optimal classification combination is achieved using the DNN + LASSO method. Our comparison

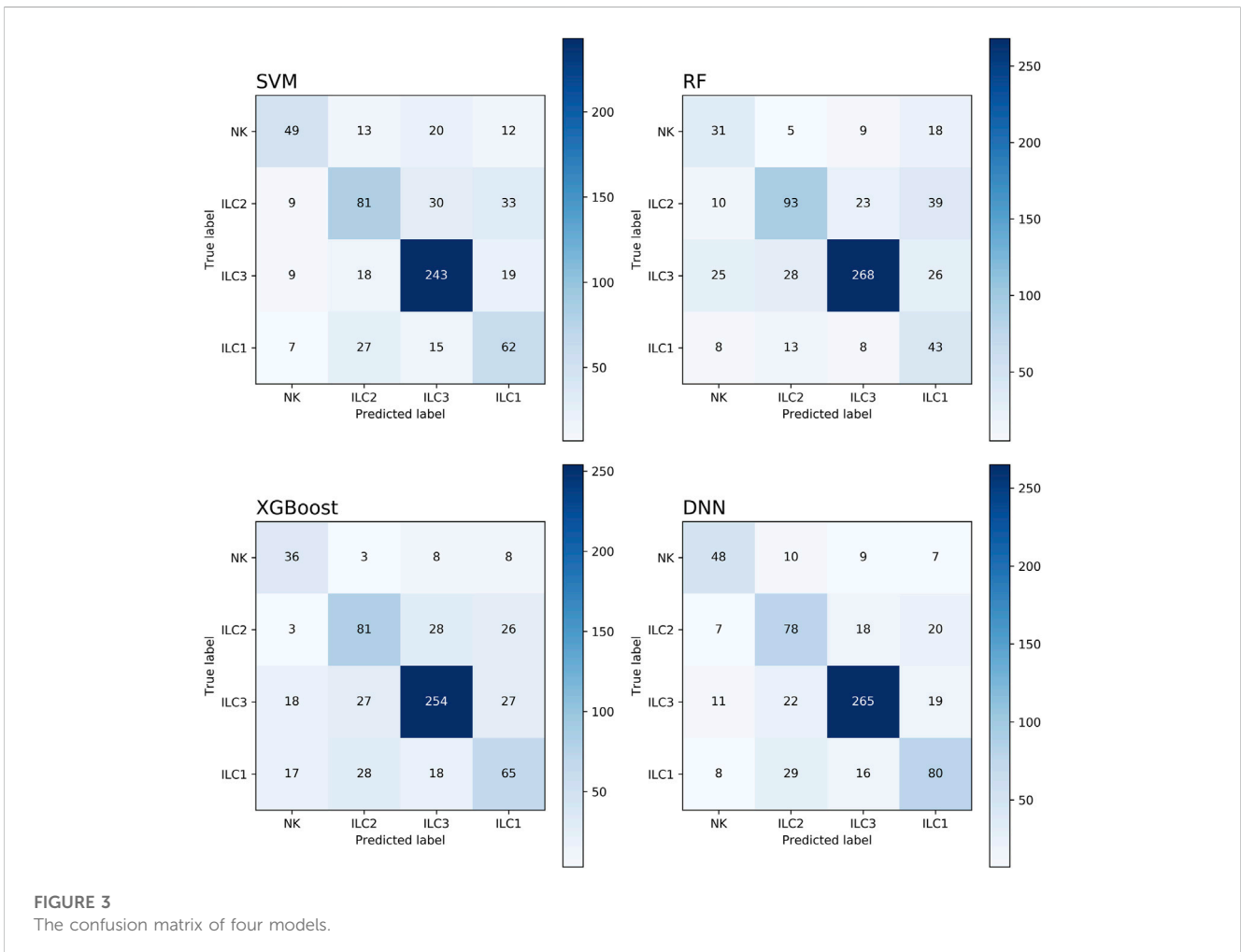


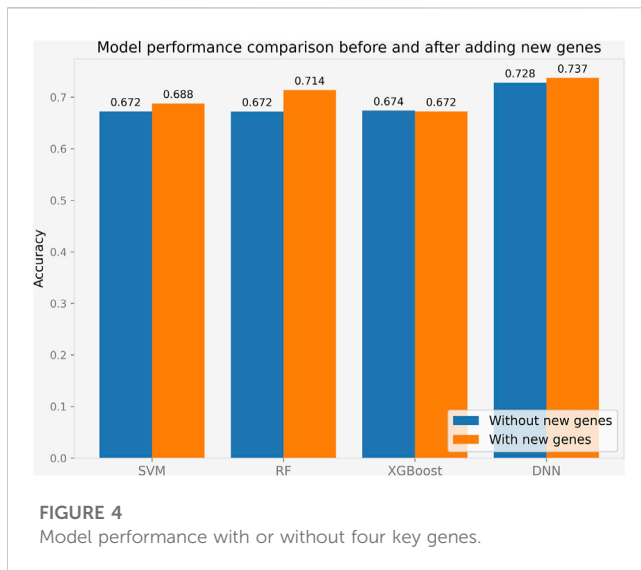
Although other feature dimensionality reduction methods seem to achieve better performance, the related gene method provides an intuitive characterization. We further verified the performance of the model on the related gene method. The confusion matrices of different machine learning classification models are presented in Figure 3, which is a widely used tool for evaluating classification model performance and measuring prediction accuracy. A confusion matrix is a two-dimensional table that represents the true and predicted labels of a classification model, where the rows correspond to the true labels and the columns correspond to the predicted labels. Each element of the matrix represents the count of samples for which the classification model predicted the category shown in the corresponding column, while the actual category was shown in the corresponding row. Based on the results shown in Figure 4, it is evident that the DNN classification model outperforms the other models. These findings suggest that the DNN model may be a more effective approach for this specific classification task.

3.3 Model performance with key genes

analysis provides valuable insights into the effectiveness of different feature selection methods and classification models for our specific dataset, and our results can inform future research in this field.

We performed a literature investigation and identified four genes, BCL11B, BATF, RUNX3, and ZBTB46, that are associated with Innate Lymphoid Cells. To investigate whether transcripts of





these genes could improve the prediction of immune typing, we incorporated them into our model. We conducted several experimental comparisons to fully evaluate the impact of these transcripts on the model's accuracy, and the results are shown in Figure 4.

Our experiments demonstrated that the addition of these transcripts significantly enhanced the accuracy of the model. Among all the methods tested, LASSO-DNN showed the best performance. These findings suggest that incorporating the transcripts of these four genes, particularly in combination with LASSO-DNN, has the potential to improve the performance of the immune typing model. Therefore, our study highlights the importance of these genes in immune typing and provides a framework for future research in this field.

4 Conclusion

Cellular immunoassay detection currently relies on sequencing technology and biological experiments. With the continuous development of medicine, immunotherapy has significantly improved the survival rate of patients with advanced cancer. Therefore, it is of utmost clinical and basic research significance to predict immune typing in advance. To achieve this, two critical steps are necessary, namely, identifying key transcripts and building effective machine learning models to accurately predict immune typing based on these transcripts.

In this study, we identified genes associated with Innate Lymphoid Cells and obtained their corresponding transcription expression levels. We used three feature extraction methods for feature dimension reduction and designed a DNN model for predicting immune typing. We compared the performance of the DNN method with several other methods and found that the combination of DNN and

LASSO provided the best classification performance. Moreover, we tested whether the information of four genes found in literature research could improve the accuracy of our model. These four genes provided valuable information and effectively enhanced the accuracy of the model in predicting immune typing.

While our method was developed specifically for the classification of ILCs, it has the potential to be applied in other contexts. In general, the machine learning model we have developed processes numerical matrices based on biological data from cells, and can produce results for any such dataset. However, one important factor to consider is the similarity between ILCs and other cell types. If other cell types share similar gene expression patterns with ILCs, our approach may also be effective for these cell types. Conversely, if the gene expression patterns are significantly different, our approach may not be suitable.

In summary, our study presents a novel method for predicting immune typing and demonstrates the accuracy of genes previously identified in literature research. Our findings contribute to the advancement of immune typing prediction and provide a framework for future research in this field.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

Author contributions

HD: Conceptualization and Methodology. HM: Visualization and Investigation. ZD and ZZ: Supervision. HY and ZW: Writing-Reviewing and Editing. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Artis, D., and Spits, H. (2015). The biology of innate lymphoid cells. *Nature* 517 (7534), 293–301. doi:10.1038/nature14189
- Badia-i Mompel, P., Vélez Santiago, J., Braunger, J., Geiss, C., Dimitrov, D., Müller-Dott, S., et al. (2022). decoupler: ensemble of computational methods to infer biological activities from omics data. *Bioinforma. Adv.* 2 (1), vba016. doi:10.1093/bioadv/vbac016
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2012). Ncbi geo: archive for functional genomics data sets—update. *Nucleic acids Res.* 41 (D1), D991–D995. doi:10.1093/nar/gks1193
- Bartemes, K. R., and Kita, H. (2021). Roles of innate lymphoid cells (ilcs) in allergic diseases: the 10-year anniversary for ilc2s. *J. Allergy Clin. Immunol.* 147 (5), 1531–1547. doi:10.1016/j.jaci.2021.03.015
- Biau, G., and Scornet, E. (2016). A random forest guided tour. *Test* 25, 197–227. doi:10.1007/s11749-016-0481-7
- Björklund, Å. K., Forkel, M., Picelli, S., Konya, V., Theorell, J., Friberg, D., et al. (2016). The heterogeneity of human cd127+ innate lymphoid cells revealed by single-cell rna sequencing. *Nat. Immunol.* 17 (4), 451–460. doi:10.1038/ni.3368
- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., et al. (2015). *Xgboost: Extreme gradient boosting*. R package version 0.4-2 1, 1–4.
- Daffertshofer, A., Lamoth, C. J., Meijer, O. G., and Beek, P. J. (2004). Pca in studying coordination and variability: a tutorial. *Clin. Biomech.* 19 (4), 415–428. doi:10.1016/j.clinbiomech.2004.01.005
- Eberl, G., Colonna, M., Di Santo, J. P., and McKenzie, A. N. (2015). Innate lymphoid cells. Innate lymphoid cells: a new paradigm in immunology. *Science* 348 (6237), aaa6566. doi:10.1126/science.aaa6566
- Efremova, M., and Teichmann, S. A. (2020). Computational methods for single-cell omics across modalities. *Nat. methods* 17 (1), 14–17. doi:10.1038/s41592-019-0692-4
- Elemam, N. M., Hannawi, S., and Maghazachi, A. A. (2017). Innate lymphoid cells (ilcs) as mediators of inflammation, release of cytokines and lytic molecules. *Toxins* 9 (12), 398. doi:10.3390/toxins9120398
- Everaere, L., Ait Yahia, S., Bouté, M., Audoussot, C., Chenivesse, C., and Tscopoulos, A. (2018). Innate lymphoid cells at the interface between obesity and asthma. *Immunology* 153 (1), 21–30. doi:10.1111/imm.12832
- Everaere, L., Ait-Yahia, S., Molendi-Coste, O., Vorng, H., Quemener, S., LeVu, P., et al. (2016). Innate lymphoid cells contribute to allergic airway disease exacerbation by obesity. *J. Allergy Clin. Immunol.* 138 (5), 1309–1318.e11. doi:10.1016/j.jaci.2016.03.019
- Golebski, K., Ros, X. R., Nagasawa, M., van Tol, S., Heesters, B. A., Aglmous, H., et al. (2019). Il-1 β , il-23, and tgf- β drive plasticity of human ilc2s towards il-17-producing ilcs in nasal inflammation. *Nat. Commun.* 10 (1), 2162. doi:10.1038/s41467-019-09883-7
- Hazenber, M. D., and Spits, H. (2014). Human innate lymphoid cells. *J. Am. Soc. Hematol.* 124 (5), 700–709. doi:10.1182/blood-2013-11-427781
- Hériché, J.-K., Alexander, S., and Ellenberg, J. (2019). Integrating imaging and omics: computational methods and challenges. *Annu. Rev. Biomed. Data Sci.* 2, 175–197. doi:10.1146/annurev-biodatasci-080917-013328
- Hu, D., Liang, K., Zhou, S., Tu, W., Liu, M., and Liu, X. (2023). scdfc: a deep fusion clustering method for single-cell rna-seq data. *Briefings Bioinforma.* 24, bbad216. doi:10.1093/bib/bbad216
- Jung, A. (2022). “Machine learning: the basics,” in *preprint: mlbook.cs.aalto.fi* (Berlin, Germany: Springer).
- Kabata, H., Motomura, Y., Kiniwa, T., Kobayashi, T., and Moro, K. (2022). “Ilcs and allergy,” in *Innate lymphoid cells* (Berlin, Germany: Springer), 75–95.
- Kaur, P., Singh, A., and Chana, I. (2021). Computational techniques and tools for omics data analysis: state-of-the-art, challenges, and future directions. *Archives Comput. Methods Eng.* 28, 4595–4631. doi:10.1007/s11831-021-09547-0
- Kogame, T., Egawa, G., Nomura, T., and Kabashima, K. (2022). Waves of layered immunity over innate lymphoid cells. *Front. Immunol.* 13, 957711. doi:10.3389/fimmu.2022.957711
- Korchagina, A. A., Shein, S. A., Koroleva, E., and Tumanov, A. V. (2023). Transcriptional control of ilc identity. *Front. Immunol.* 14, 1146077. doi:10.3389/fimmu.2023.1146077
- Kumar, V. (2022). “Innate lymphoid cells in autoimmune diseases,” in *Translational autoimmunity* (Amsterdam, Netherlands: Elsevier), 143–175.
- Li, F., Zhou, Y., Zhang, X., Tang, J., Yang, Q., Zhang, Y., et al. (2020). Ssizer: determining the sample sufficiency for comparative biological study. *J. Mol. Biol.* 432 (11), 3411–3421. doi:10.1016/j.jmb.2020.01.027
- Li, S., Yang, D., Peng, T., Wu, Y., Tian, Z., and Ni, B. (2017). Innate lymphoid cell-derived cytokines in autoimmune diseases. *J. Autoimmun.* 83, 62–72. doi:10.1016/j.jaut.2017.05.001
- Miller, D., Motomura, K., Garcia-Flores, V., Romero, R., and Gomez-Lopez, N. (2018). Innate lymphoid cells in the maternal and fetal compartments. *Front. Immunol.* 9, 2396. doi:10.3389/fimmu.2018.02396
- Montaldo, E., Juelke, K., and Romagnani, C. (2015). Group 3 innate lymphoid cells (ilc3s): origin, differentiation, and plasticity in humans and mice. *Eur. J. Immunol.* 45 (8), 2171–2182. doi:10.1002/eji.201545598
- Osborne, M. R., Presnell, B., and Turlach, B. A. (2000). On the lasso and its dual. *J. Comput. Graph. statistics* 9 (2), 319–337. doi:10.1080/10618600.2000.10474883
- Piñero, J., Ramírez-Anguita, J. M., Saüch-Pitarch, J., Ronzano, F., Centeno, E., Sanz, F., et al. (2020). The disgenet knowledge platform for disease genomics: 2019 update. *Nucleic acids Res.* 48 (D1), D845–D855. doi:10.1093/nar/gkz1021
- Surace, L., and Wilhelm, C. (2022). Keeping ilcs in shape: pd-1 as a metabolic checkpoint. *Nat. Metab.* 4 (7), 794–795. doi:10.1038/s42255-022-00599-5
- Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., et al. (2016). The perseus computational platform for comprehensive analysis of (prote) omics data. *Nat. methods* 13 (9), 731–740. doi:10.1038/nmeth.3901
- Vacca, P., Chiossone, L., Mingari, M. C., and Moretta, L. (2019). Heterogeneity of nk cells and other innate lymphoid cells in human and murine decidua. *Front. Immunol.* 10, 170. doi:10.3389/fimmu.2019.00170
- Vivier, E., Artis, D., Colonna, M., Diefenbach, A., Di Santo, J. P., Eberl, G., et al. (2018). Innate lymphoid cells: 10 years on. *Cell* 174 (5), 1054–1066. doi:10.1016/j.cell.2018.07.017
- Wang, H., and Hu, D. (2005). “Comparison of svm and ls-svm for regression,” in 2005 International conference on neural networks and brain, Hilton, July 31 - August 4, 2005, 279–283.
- Watson, E. R., Taherian Fard, A., and Mar, J. C. (2022). Computational methods for single-cell imaging and omics data integration. *Front. Mol. Biosci.* 8, 768106. doi:10.3389/fmolb.2021.768106