



A Multi-task Method for Immunofixation Electrophoresis Image Classification

Yi Shi¹, Rui-Xiang Li¹, Wen-Qi Shao², Xin-Cen Duan², Han-Jia Ye¹(✉),
De-Chuan Zhan¹, Bai-Shen Pan², Bei-Li Wang², Wei Guo², and Yuan Jiang¹

¹ State Key Laboratory for Novel Software Technology, Nanjing University,
Nanjing, China

yehj@lamda.nju.edu.cn

² Department of Laboratory Medicine, Zhongshan Hospital, Fudan University,
Shanghai, China

Abstract. In the field of plasma cell disorders diagnosis, the detection of abnormal monoclonal (M) proteins through Immunofixation Electrophoresis (IFE) is a widely accepted practice. However, the classification of IFE images into nine distinct categories is a complex task due to the significant class imbalance problem. To address this challenge, a two-sub-task classification approach is proposed, which divides the classification task into the determination of severe and mild cases, followed by their combination to produce the final result. This strategy is based on the expert understanding that the nine classes are different combinations of severe and mild cases. Additionally, the examination of the dense band co-location on the electrophoresis lane and other lanes is crucial in the expert evaluation of the image class. To incorporate this expert knowledge into the model training, inner-task and inter-task regularization is introduced. The effectiveness of the proposed method is demonstrated through experiments conducted on approximately 15,000 IFE images, resulting in interpretable visualization outcomes that are in alignment with expert expectations. Codes are available at <https://github.com/shiy19/IFE-classification>.

Keywords: IFE image classification · Multi-task learning ·
Task-related regularization · Class imbalance · Expert knowledge

1 Introduction

The utilization of immunofixation electrophoresis (IFE) as a laboratory technique has been widely adopted for the identification and characterization of abnormal proteins in blood or urine specimens [11, 15, 25]. This technique is commonly employed to detect monoclonal gammopathy, a condition characterized

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43987-2_15.

by the presence of deviant monoclonal (M) proteins, which can aid in the diagnosis of multiple myeloma, Waldenstrom’s macroglobulinemia, and other plasma cell disorders [2, 16, 18, 26]. The electrophoresis process separates serum proteins into distinct lanes, which are then treated with specific antisera against ‘IgG’, ‘IgA’, ‘IgM’, ‘ κ ’, and ‘ λ ’. Precipitin bands indicative of abnormal M-proteins exhibit greater density and darkness than their normal counterparts. The electrophoresis (ELP) lane, which represents a mixture of multiple proteins, serves as a reference for the recognition of various M-proteins. For human experts, identification of an M-protein is achieved by observing the co-location of bands between the ELP lane, heavy chain lanes (‘G’, ‘A’, and ‘M’), and light chain lanes (‘ κ ’ and ‘ λ ’). For example, as shown in the top of Fig. 1(a), if the dense (dark) band of the ELP lane can be aligned with that of the heavy chain lane ‘G’ and the light chain lane ‘ κ ’, this sample can be identified as ‘IgG- κ positive’. In the same way, we can identify the other two samples in Fig. 1(a) as ‘ λ positive’ and ‘Negative’.

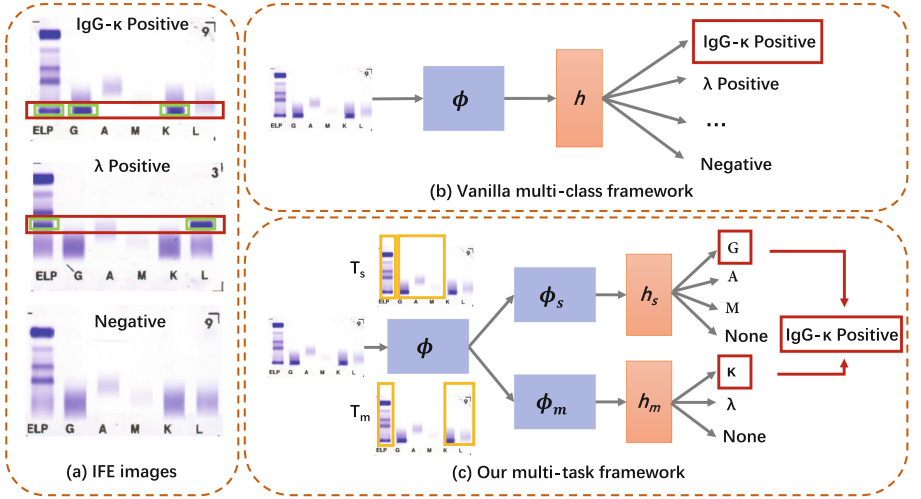


Fig. 1. (a) Three kinds of IFE images: ‘IgG- κ positive’ (severe case and mild case simultaneously), ‘ λ positive’ (only mild case), and ‘Negative’. The area in the red box is the judgment area marked by our experts, and the area in the green box is the dense band. (b) An illustration of vanilla multi-class framework, which directly performs a 9-class classification for IFE images. (c) An illustration of our proposed multi-task framework, which splits the task into two sub-tasks: T_s and T_m to deal with the classification of severe cases and mild cases, respectively. T_s and T_m can focus on corresponding areas, as shown in the orange boxes. ϕ and h represent feature extractors and classifiers, respectively. (Color figure online)

Recently, the application of machine learning for image recognition has gained attention as a means of handling large-scale samples and providing objective

and quality results [19, 20, 23, 28]. While machine learning has been successfully applied to a wide range of medical images, such as Computed Tomography (CT) [4, 13], histopathological images [8, 31], magnetic resonance images [3, 12], ultrasound images [17, 29], and mammograms [21, 30], its application to IFE presents unique challenges. IFE images encompass nine distinct yet interrelated classes, including ‘Negative’, two types of positive with mild cases (‘ κ ’, ‘ λ ’), and six types of positive with both severe and mild cases (‘IgG- κ ’, ‘IgG- λ ’, ‘IgA- κ ’, ‘IgA- λ ’, ‘IgM- κ ’, ‘IgM- λ ’)¹, which can be identified as different combinations of three severe cases (‘IgG’, ‘IgA’, and ‘IgM’) and two mild cases (‘ κ ’ and ‘ λ ’). The class imbalance phenomenon [1, 9] is common in IFE images, as demonstrated by our dataset collected from a large hospital, where the predominant class ‘Negative’ comprises over 10,000 samples, while the minor class ‘ κ positive’ has only 63 samples. Conventional multi-class classification approaches struggle to accurately learn the features of minor classes, leading to subpar performance, particularly for minor classes. Previous studies have approached IFE image classification as a multi-class problem, such as the two-stage strategy proposed by Hu et al. [6] which first classifies samples into positive and negative categories, followed by classification into different positive categories. Wei et al. [24] employed a dynamic programming based method to segment IFE images into lanes and strips, calculating the similarity between them as inputs to the model. However, these approaches necessitate additional processing or complex data preprocessing, and they overlook the class imbalance problem.

Inspired by the expert knowledge that all the nine classes are different combinations of severe cases and mild cases, we propose a novel end-to-end multi-task learning method for IFE image classification, as depicted in Fig. 1(c). Our approach employs a decompose-and-fuse strategy, where the classification task is first divided into two sub-tasks: one for the classification of severe cases, and the other for mild cases. This decomposition effectively mitigates the class imbalance present in the original multi-class problem. As shown in Fig. 2(a), while ‘IgG- κ positive’, ‘IgA- κ positive’, ‘IgM- κ positive’, and ‘ κ positive’ are considered minor classes compared to class ‘Negative’ in the original task, they all contribute to ‘ κ ’. Thus, in the sub-task that classifies mild cases, class ‘ κ ’ has a higher number of samples, reducing the class imbalance issue. The same holds for the sub-task that categorizes severe cases. Due to the reduced class imbalance in the two sub-tasks, improved performance can be achieved. The results from the two sub-tasks are then combined to produce good performance in the original 9-class classification problem. Additionally, we have integrated an attention module into the convolution layers at the height level of the image to enhance the model’s focus on relevant image features and improve visualization and explanation. Drawing upon expert knowledge, we have implemented two forms of regularization to more accurately model the relationships within and between tasks, as depicted in Fig. 2(b, c). The inner-task regularization smoothens the attention map of negative samples and sharpens that of positive samples, which aligns with the expert

¹ Samples with multiple positives are rare, *e.g.*, with ‘IgG- κ ’ and ‘IgM- λ ’ simultaneously. These samples are not considered in this study.

understanding that only positive samples exhibit co-located dense bands. For a sample with both severe and mild cases simultaneously (namely S&M-sample), the inter-task regularization reduces the gap between its attention maps from the two sub-tasks, which corresponds to the expert knowledge that dense bands co-exist on both the severe lane and the mild lane for a S&M-sample.

In summary, our contributions are three-fold: (1) we propose an end-to-end multi-task method for IFE image classification that effectively addresses the class imbalance issue, (2) we implement two types of regularization mechanisms to more accurately model relationships within and between tasks, and (3) our extensive experiments and visualization demonstrate the effectiveness and explainability of our proposed method.

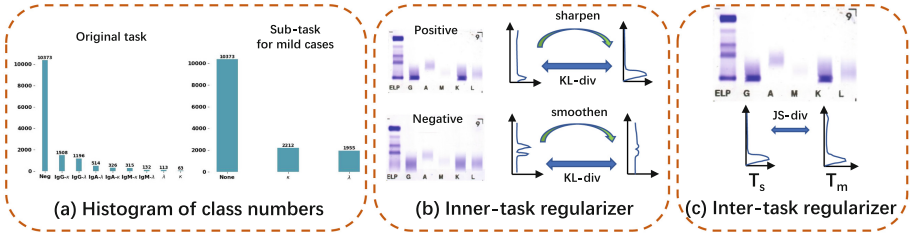


Fig. 2. (a) Histogram of class numbers for the original multi-class task and our sub-task that classifies mild cases. (b) An illustration of our proposed inner-task regularizer. We sharpen/smoothen the attention map of positive/negative samples first, and minimize its KL-divergence with the original attention map. (c) An illustration of our proposed inter-task regularizer. For positive samples with severe and mild case simultaneously, we minimize the JS-divergence between the attention maps of sub-task T_s and sub-task T_m .

2 Method

In this section, we describe the methodology for the development of a novel multi-task learning approach for the classification of IFE images. Based on the two-sub-task framework, our approach employs both within-task and between-task regularization on the attention module to address several critical challenges, such as alleviating class imbalance and modeling different patterns of severe and mild cases, which can incorporate expert knowledge into the model training.

2.1 Multi-task Framework

Directly addressing the classification of IFE images as a multi-class (9-class) problem may face a severe class imbalance phenomenon, which makes the classification of IFE images a challenging task. Inspired by expert knowledge, we decompose the multi-class problem into two sub-tasks to mitigate this issue.

The first sub-task is to classify heavy lanes into four classes ('None', 'G', 'A', and 'M'), while the second is to classify light lanes into three classes ('None', ' κ ', and ' λ '). If a sample does not possess the corresponding dense bands in its heavy or light lane that match with the ELP lane, it is classified as 'None'. This decomposition approach ensures that different minor classes of the original 9-class task that shares the same severe or mild cases are combined into larger minor classes for the two sub-tasks, which can be proven by the histogram in Fig. 2(a). Define imbalance ratio (IR) as follows:

$$IR = N_{\min}/N_{\max} \quad (1)$$

Here, N_{\min} and N_{\max} represent minority class size and majority class size, respectively. In reviewing our dataset, the initial task demonstrates an IR of 0.06, whereas the two sub-tasks exhibit an IR of 0.188 and 0.042, respectively. Utilizing the metric of IR provides an academically rigorous, quantitative methodology to discern the mitigation of the class imbalance issue within the given dataset. The final classification result is obtained by combining the prediction results from two sub-tasks. As demonstrated in Fig. 1(c), when two sub-tasks yield predictions of 'G' and ' κ ', the final output is identified as 'IgG- κ positive'. In each sub-task, the ResNet18 architecture [5] is employed as the feature extractor for raw images, while a fully connected layer serves as the classifier. To capture the relationship between the two sub-tasks, and to ensure that they both learn the basic features of the raw image, such as the dense band in the lane, the first three blocks of the ResNet18 model are shared between the two sub-tasks. This sharing of blocks was found to provide the best performance empirically, and further details can be found in the supplementary material. Formally speaking, for an instance $\mathbf{x}_i \in \mathbb{R}^D$, we split its ground-truth label $\mathbf{y}_i \in \mathbb{S}(9)$ into $\mathbf{y}_{is} \in \mathbb{S}(4)$ and $\mathbf{y}_{im} \in \mathbb{S}(3)$. Here, $\mathbb{S}(k)$ means the set of $\{1, 2, \dots, k\}$, and subscripts s and m represent the task that classifies severe cases and mild cases, respectively. Let $\phi(\cdot)$ and $\mathbf{h}(\cdot)$ represent the feature extractor and the classifier, respectively. The logit vectors of a sample \mathbf{x}_i in two sub-tasks are obtained with $\hat{\mathbf{f}}_{is} = \mathbf{h}_s(\phi_s(\mathbf{x}_i))$ and $\hat{\mathbf{f}}_{im} = \mathbf{h}_m(\phi_m(\mathbf{x}_i))$. The predicted label for each sub-task is then obtained by taking the maximum value from the corresponding logit vector. ϕ_s and \mathbf{h}_s (ϕ_m and \mathbf{h}_m) of the sub-task are trained with cross-entropy loss \mathbf{L}_{CE}^s (\mathbf{L}_{CE}^m), which minimizes the discrepancy between the posterior class probability with the ground-truth.

2.2 Task-Related Regularization

Horizontal Attention. As depicted in Fig. 1(a), human experts identify the valid part of an IFE image as a horizontal rectangular region that displays the alignment of dense bands between the ELP lane and the other lanes. To encourage the model to focus on this area, an attention mechanism has been integrated into the convolutional layer of the proposed method. Attention mechanisms [7, 22, 27] are widely used in Convolutional Neural Network (CNN) models

to focus on specific channels or spatial regions. However, for IFE image classification, only attention to the horizontal direction is required. The horizontal attention mechanism used in this study differs from conventional spatial attention in that it only focuses on a 1D direction (the horizontal direction). To calculate the horizontal attention, average-pooling $\text{AvgPool}(\mathbf{F})$ and max-pooling $\text{MaxPool}(\mathbf{F})$ are applied along the channel axis and the width axis, followed by a standard convolution layer and a sigmoid function σ to generate a horizontal attention map $\mathbf{M}(\mathbf{F}) \in \mathbb{R}^H: \mathbf{M}(\mathbf{F}) = \sigma(\text{Conv}([\text{AvgPool}(\mathbf{F}); \text{MaxPool}(\mathbf{F})]))$. Here, \mathbf{F} represents the feature vector, and H is the height of the image. Compared to spatial attention, this attention map encodes the horizontal region to be emphasized, enabling the model to better capture the relationship between the dense bands in different lanes. Furthermore, the horizontal attention mechanism provides a convenient way to visualize the model results, making it more interpretable, as illustrated in Fig. 3.

Inner-Task Regularization. To effectively model the patterns in different samples, it is crucial to understand the desired appearance of the attention map for positive and negative samples. As previously discussed, for positive samples, the model should emphasize the horizontal region that displays the alignment of dense bands. However, obtaining human-labeled attention areas for the training phase is a time-consuming and labor-intensive task. To address this issue, the attention map is differentiated between positive and negative samples. The goal is to sharpen the attention map for positive samples and smoothen it for negative samples. The attention map $\mathbf{M}(\mathbf{F})$ can be viewed as the distribution of attention intensity along the height dimension of the image, as depicted in Fig. 2(b). The distribution is adjusted to be sharper or flatter by applying a Softmax operator with a temperature parameter τ : $\bar{\mathbf{M}}(\mathbf{F}) = \text{Softmax}(\mathbf{M}(\mathbf{F})/\tau)$. The inner-task regularization $\text{Reg}_{\text{inner}}^s$ and $\text{Reg}_{\text{inner}}^m$ can be designed using the Kullback-Leibler divergence (KLdiv) [10] as follows:

$$\begin{aligned} \text{Reg}_{\text{inner}}^s &= \text{KLdiv}(\bar{\mathbf{M}}_s(\mathbf{F}_s) \cdot \mathbf{M}_s(\mathbf{F}_s)) \\ \text{Reg}_{\text{inner}}^m &= \text{KLdiv}(\bar{\mathbf{M}}_m(\mathbf{F}_m) \cdot \mathbf{M}_m(\mathbf{F}_m)) \end{aligned} \quad (2)$$

Here, $\text{KLdiv}(A \cdot B) = A \cdot (\log A - \log B)$. The inner-task attention mechanism allows the model to differentiate the attention map between positive and negative samples.

Inter-Task Regularization. It is important to note that the alignment of dense bands between heavy lanes, light lanes, and the ELP lane is a crucial factor in the identification of S&M-samples by human experts. To model this relationship in the proposed method, the two sub-tasks are encouraged to emphasize similar horizontal areas. Specifically, for all S&M-samples, the inter-task regularization can be designed with the Jensen-Shannon Divergence (JSdiv) [14] as follows :

$$\text{Reg}_{\text{inter}} = \text{JSdiv}(\mathbf{M}_s(\mathbf{F}_s) \cdot \mathbf{M}_m(\mathbf{F}_m)) \quad (3)$$

Here, $\text{JSdiv}(A \cdot B) = (\text{KLdiv}(A \cdot C) + \text{KLdiv}(B \cdot C))/2$, and $C = (A + B)/2$. The inter-task regularizer enables the model to better capture the patterns of S&M-samples by modeling the relationship between the two sub-tasks.

2.3 Overall Objective

The overall optimization objective for the proposed method can be expressed as:

$$\mathbf{L}_{\text{CE}}^{\text{s}} + \mathbf{L}_{\text{CE}}^{\text{m}} + \lambda_{\text{inner}} \cdot (\mathbf{Reg}_{\text{inner}}^{\text{s}} + \mathbf{Reg}_{\text{inner}}^{\text{m}}) + \lambda_{\text{inter}} \cdot \mathbf{Reg}_{\text{inter}}. \quad (4)$$

Here, λ_{inner} and λ_{inter} are hyper-parameters.

In conclusion, the proposed multi-task method mitigates the class imbalance problem in the classification of IFE images by decomposing the task into two separate sub-tasks. The use of inner-task and inter-task regularization on the horizontal attention mechanism allows the model to capture the relationships between sub-tasks and learn discriminative patterns for both positive and negative samples.

3 Experiment

Dataset. We introduce a new IFE image dataset, which comprises 14540 images along with their respective protein presence class labels. As defined in Eq. 1, the dataset exhibits an imbalance ratio of 0.006, which underscores a pronounced instance of class imbalance within the dataset. Our dataset is collected from a large hospital between January 2020 and December 2022, and it reflects the real-world environment, as the images usually contain certain types of noise, such as left and right offset of lanes and non-pure white background color. Unlike previous works [6, 24] that perform preprocessing operations, such as cutting the images into different horizontal or vertical pieces, the raw images are utilized in the proposed method, resulting in an end-to-end approach. The training and test sets are stratified based on a 4:1 ratio, and the dataset characteristics can be found in the supplementary material. The images are resized to 224×224 as the input for the model.

Table 1. Performance comparison on IFE dataset. The best performances are in bold.

Methods	Acc (%)	F1-score(%)	Fnr (%)	M1-Acc (%)	M3-Acc (%)
Multi-class	95.12 \pm 0.04	85.16 \pm 0.60	10.10 \pm 0.41	88.46 \pm 3.84	73.83 \pm 2.04
Collocative Net [24]	93.07 \pm 0.05	81.80 \pm 0.08	14.22 \pm 0.48	80.77 \pm 3.84	59.97 \pm 0.65
Two stage DNN [6]	92.54 \pm 0.63	75.68 \pm 0.96	9.31 \pm 0.60	42.30 \pm 3.85	42.03 \pm 0.64
Our Method	95.43 \pm 0.10	86.68 \pm 0.24	8.54 \pm 0.18	92.31 \pm 0.00	76.10 \pm 1.28

Table 2. Ablation study on the multi-task framework and two regularization.

Multi-task	Reg _{inner}	Reg _{inter}	Acc (%)	F1-score (%)	Fnr (%)	M1-Acc (%)	M3-Acc (%)
\times	\times	\times	95.12 \pm 0.04	85.16 \pm 0.60	10.10 \pm 0.41	88.46 \pm 3.84	73.83 \pm 2.04
\checkmark	\times	\times	95.23 \pm 0.01	85.79 \pm 0.35	9.80 \pm 0.12	92.31 \pm 0.00	73.60 \pm 0.46
\checkmark	\checkmark	\times	95.37 \pm 0.05	86.16 \pm 0.15	9.14 \pm 0.42	92.31 \pm 0.00	75.99 \pm 1.39
\checkmark	\times	\checkmark	95.24 \pm 0.08	85.70 \pm 0.85	9.26 \pm 0.06	88.46 \pm 3.84	75.00 \pm 0.01
\checkmark	\checkmark	\checkmark	95.43 \pm 0.10	86.68 \pm 0.24	8.54 \pm 0.18	92.31 \pm 0.00	76.10 \pm 1.28

Implementation Details. All models are implemented using the PyTorch framework and trained on an NVIDIA GeForce RTX 2080ti GPU with 11 GB memory. The network is trained for 100 epochs with a batch size of 64, and the optimization is performed using Stochastic Gradient Descent (SGD) with momentum as the optimizer. The initial learning rate is set to 0.03 for the first 5 epochs and then cosine-decayed for the remaining epochs. The training stage takes 1.2 h on one GPU and occupies 4G display memory. The hyper-parameters of regularization are tuned based on cross-validation on the training set, with the best hyper-parameters set to $\lambda_{\text{inner}} = 0.3$, $\tau = 5$, $\lambda_{\text{inter}} = 0.3$.

Evaluation Metric and Comparison Methods. We assess the efficacy via a suite of evaluation metrics, including accuracy (acc), F1-score, false negative rate (fnr), top-1 minor class accuracy (M1-Acc), and top-3 minor class accuracy (M3-Acc). These metrics are widely adopted for classification problems, with the false negative rate being of particular significance in the medical domain as it quantifies the likelihood of misidentifying positive samples as negative. The top-1 and top-3 minor class accuracy metrics evaluate the model’s performance on underrepresented classes. A ResNet18-based multi-class classification approach is selected as the baseline and tuned using an identical range of hyper-parameters as the proposed method. The baseline is trained for 100 epochs with a batch size of 64, an initial learning rate of 0.1, and a decay scheduler consistent with the proposed method. Additionally, the proposed method is compared against Two Stage DNN [6] and Collocative Net [24], the state-of-the-art methods in the IFE recognition domain, using our dataset. The results are obtained by adhering to the code released by the authors. The dataset is preprocessed manually prior to network training in accordance with the methodology described in the respective papers.

Performance Evaluation. The methodology proposed in this study is thoroughly evaluated and compared to both the baseline method and other state-of-the-art techniques on the IFE dataset, as presented in Table 1. The results reveal that our method outperforms all other methods in terms of all evaluation metrics, particularly with respect to top1 and top3 minor class accuracy, thus highlighting its efficacy in addressing the class imbalance problem and achieving improved performance for underrepresented classes. Our empirical analysis demonstrates that Collocative Net [24] and Two Stage DNN [6] exhibit inferior performance compared to the conventional multi-class approach. This can be attributed to the excessive manual intervention employed during preprocessing and model development, which is not suitable for handling real-world IFE datasets containing various forms of noise, such as lane offsets and impurity pixels in the background.

Visualization. The proposed horizontal attention mechanism provides a useful tool for verifying the consistency of model-learned knowledge with expert experience. As illustrated in Fig. 3, the areas with the highest attention intensity are highlighted in orange, while red boxes indicate the judgment areas of our experts. The results show that the attention regions of positive samples largely

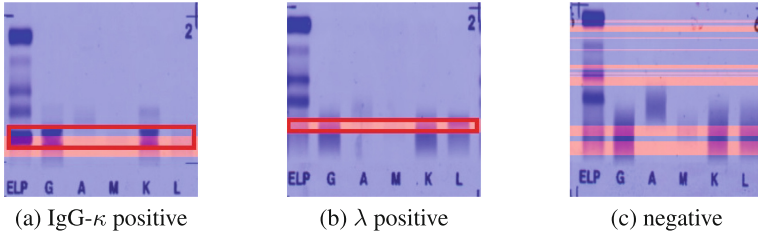


Fig. 3. Visualization of horizontal attention maps on different types of IFE images. The orange areas are the most attention part learned by our model, and the areas in red boxes are the judgment areas marked by our experts. (Color figure online)

overlap with the expert judgment areas, while the attention regions of negative samples are more dispersed, indicating a lack of specific areas of interest. These findings demonstrate that the knowledge learned by our model is consistent with expert experience.

Ablation Study. An ablation study is conducted to examine the contribution of key components of the proposed model, including the multi-task framework and two regularization methods. The results, presented in Table 2, indicate that the use of the multi-task framework results in significant performance gains, particularly for minority classes, and that the introduction of the two regularization methods further improves the overall performance of the model. Further exploration of the effects of different hyper-parameters can be found in the supplementary materials.

4 Conclusion

In this study, we investigate the issue of classifying immunofixation electrophoresis (IFE) images and contend that conventional multi-class classification strategies may not be adequate due to the considerable class imbalance issue. To address this issue, we introduce a multi-task framework that leverages expert knowledge. Our method effectively captures different patterns of both severe and mild cases through the use of inner-task and inter-task regularization. The empirical evaluation of our method, conducted on a dataset of approximately 15,000 IFE images, demonstrates its superiority over other approaches. The results also provide interpretable visualizations that align with expert annotations and highlight the potential for improved accuracy and reliability in the classification of IFE images. This development has the potential to significantly advance the diagnostic capabilities of monoclonal protein disorders in the medical field.

Acknowledgments. This research was supported by National Key R&D Program of China (2020AAA0109401), NSFC (62176117, 61921006, 62006112), Collaborative Innovation Center of Novel Software Technology and Industrialization.

References

1. Abd Elrahman, S.M., Abraham, A.: A review of class imbalance problem. *J. Network Innovative Comput.* **1**(2013), 332–340 (2013)
2. Cowan, A.J., et al.: Diagnosis and management of multiple myeloma: a review. *JAMA* **327**(5), 464–477 (2022)
3. Dong, S., et al.: Multi-scale super-resolution magnetic resonance spectroscopic imaging with adjustable sharpness. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) *MICCAI 2022. LNCS*, vol. 13436, pp. 410–420. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16446-0_39
4. Gu, Y., et al.: BMD-GAN: bone mineral density estimation using x-ray image decomposition into projections of bone-segmented quantitative computed tomography using hierarchical learning. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) *MICCAI 2022. LNCS*, vol. 13436, pp. 644–654. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16446-0_61
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *CVPR*, pp. 770–778 (2016)
6. Hu, H., et al.: Expert-level immunofixation electrophoresis image recognition based on explainable and generalizable deep learning. *Clin. Chem.* **69**(2), 130–139 (2023)
7. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: *CVPR*, pp. 7132–7141 (2018)
8. Jimenez, G., et al.: Visual deep learning-based explanation for Neuritic plaques segmentation in Alzheimer’s disease using weakly annotated whole slide histopathological images. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) *MICCAI 2022. LNCS*, vol. 13432, pp. 336–344. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16434-7_33
9. Johnson, J.M., Khoshgoftaar, T.M.: Survey on deep learning with class imbalance. *J. Big Data* **6**(1), 1–54 (2019). <https://doi.org/10.1186/s40537-019-0192-5>
10. Joyce, J.M.: Kullback-leibler divergence. In: *International Encyclopedia of Statistical Science*, pp. 720–722 (2011)
11. Keren, D.F.: *High-Resolution Electrophoresis and Immunofixation: Techniques and Interpretation* (2017)
12. Kuzmina, E., Razumov, A., Rogov, O.Y., Adalsteinsson, E., White, J., Dylov, D.V.: Autofocusing+: noise-resilient motion correction in magnetic resonance imaging. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) *MICCAI 2022. LNCS*, vol. 13436, pp. 365–375. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16446-0_35
13. López Diez, P., et al.: Deep reinforcement learning for detection of inner ear abnormal anatomy in computed tomography. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) *MICCAI 2022. LNCS*, vol. 13433, pp. 697–706. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16437-8_67
14. Menéndez, M., Pardo, J., Pardo, L., Pardo, M.: The jensen-shannon divergence. *J. Franklin Inst.* **334**(2), 307–318 (1997)
15. Moore, A.R., Avery, P.R.: Protein characterization using electrophoresis and immunofixation; a case-based review of dogs and cats. *Veterinary Clin. Pathol.* **48**, 29–44 (2019)
16. Moreau, P., et al.: Treatment of relapsed and refractory multiple myeloma: recommendations from the international myeloma working group. *Lancet Oncol.* **22**(3), e105–e118 (2021)

17. Oh, S., Kim, M.G., Kim, Y., Jung, G., Kwon, H., Bae, H.M.: Sensor geometry generalization to untrained conditions in quantitative ultrasound imaging. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13436, pp. 780–789. Springer, Cham (2022). <https://doi.org/10.1007/978-3-031-16446-0-74>
18. Rajkumar, S.V., Kumar, S.: Multiple myeloma current treatment algorithms. *Blood Cancer J.* **10**(9), 94 (2020)
19. Shin, Y., et al.: Digestive organ recognition in video capsule endoscopy based on temporal segmentation network. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13437, pp. 136–146. Springer, Cham (2022). <https://doi.org/10.1007/978-3-031-16449-1-14>
20. Shui, Z., et al.: End-to-end cell recognition by point annotation. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13434, pp. 109–118. Springer, Cham (2022). <https://doi.org/10.1007/978-3-031-16440-8-11>
21. Tardy, M., Mateus, D.: Leveraging multi-task learning to cope with poor and missing labels of mammograms. *Front. Radiol.* **1**, 796078 (2022)
22. Wang, X., Girshick, R., Gupta, A., He, K.: Non-local neural networks. In: CVPR, pp. 7794–7803 (2018)
23. Wang, Y., et al.: Key-frame guided network for thyroid nodule recognition using ultrasound videos. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13434, pp. 238–247. Springer, Cham (2022). <https://doi.org/10.1007/978-3-031-16440-8-23>
24. Wei, X.Y., et al.: Deep collocative learning for immunofixation electrophoresis image analysis. *IEEE Trans. Med. Imaging* **40**(7), 1898–1910 (2021)
25. Wilhite, D., Arfa, A., Cotter, T., Savage, N.M., Bollag, R.J., Singh, G.: Multiple myeloma: detection of free monoclonal light chains by modified immunofixation electrophoresis with antisera against free light chains. *Pract. Lab. Med.* **27**, e00256 (2021)
26. Willrich, M.A., Murray, D.L., Kyle, R.A.: Laboratory testing for monoclonal gammopathies: focus on monoclonal gammopathy of undetermined significance and smoldering multiple myeloma. *Clin. Biochem.* **51**, 38–47 (2018)
27. Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: CBAM: convolutional block attention module. In: ECCV, pp. 3–19 (2018)
28. Xia, C., Wang, J., Qin, Y., Gu, Y., Chen, B., Yang, J.: An end-to-end combinatorial optimization method for r-band chromosome recognition with grouping guided attention. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2023. LNCS, vol. 13434, pp. 3–13. Springer, Cham (2022). <https://doi.org/10.1007/978-3-031-16440-8-1>
29. Xie, Y., Liao, H., Zhang, D., Chen, F.: Uncertainty-aware cascade network for ultrasound image segmentation with ambiguous boundary. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13434, pp. 268–278. Springer, Cham (2022). <https://doi.org/10.1007/978-3-031-16440-8-26>
30. Yu, X., Pang, W., Xu, Q., Liang, M.: Mammographic image classification with deep fusion learning. *Sci. Rep.* **10**(1), 14361 (2020)
31. Zeng, B., et al.: Semi-supervised PR virtual staining for breast histopathological images. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13432, pp. 232–241. Springer, Cham (2022)