ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

Achieving a New Artificial Intelligence System for Serum Protein Electrophoresis to Recognize M-Spikes

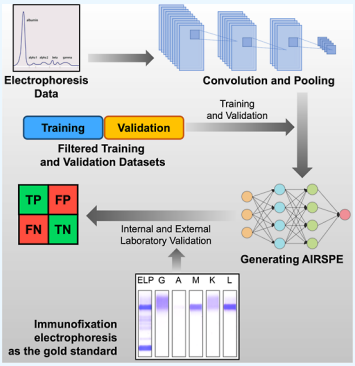
Ruojian Shen, Yuyi Hu, Qun Wu, Jing Zhu, Wen Xu, Baishen Pan, Wenqi Shao,* Beili Wang,* and Wei Guo*

Cite This: *ACS Omega* 2025, 10, 5770−5777

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Purpose: In order to accurately identify the low-concentration M-spikes in serum protein electrophoresis (SPE) patterns, a new artificial intelligence (AI) system is explored. Methods: 166,003 SPE data sets, which were equally divided into 4 training sets and 1 optimal set, were utilized to establish and evaluate the AI system, namely, "AIRSPE". 10,014 internal test sets and 1861 external test sets with immunofixation electrophoresis (IFE) results as gold standard were used to assess the performance of AIRSPE including sensitivity, negative predictive value, and concordance. In the internal test group with different concentrations of M-spikes, the consistencies of AIRSPE and manual interpretation with IF-positive results were compared. Results: AIRSPE selected MobileNetv2, which performed with F1-score of 84.60%, precision of 76.20%, recall of 95.20%, loss of 26.80%, accuracy of 89.48%, and interpretation time of 14 ms. In internal test sets, the sensitivity and negative predictive values of AIRSPE were 95.21% and 97.65%, respectively, with no significant difference in performance compared to the external test set ($P > 0.05$). AIRSPE and IFE results showed a concordance ($k = 0.832$) that implies an almost perfect agreement, which was higher than that between manual interpretation and IFE ($k = 0.699$). The M-spikes that were identified by IFE as positive in the internal test data sets that were detected by AIRSPE but were not detected by manual interpretation were mainly concentrated in the γ-fraction, with M-spike concentrations lower than 0.5 g/L. Conclusions: AIRSPE, established through AI deep learning and validated by IFE results, significantly outperforms manual interpretation in detecting low-concentration M-spikes, demonstrating its potential to assist with clinical screening for M-spikes.



1. INTRODUCTION

Monoclonal gammopathies represent a spectrum of disorders with varying manifestations ranging from a completely asymptomatic condition picked up on laboratory testing to a catastrophic, life-threatening presentation in the clinic. The use of novel therapy together with autologous stem cell transplantation lengthened median overall survival to 8 to 10 years.[1]

Serum protein electrophoresis (SPE) is commonly used as an adjunct diagnostic tool for monoclonal gammopathy.[2] Capillary zone electrophoresis (CZE), which is commonly used in clinical practice, is a cost-effective, convenient, and rapid screening method for monoclonal gammopathy.[3,4]

IFE can be used for the qualitative detection of various protein levels. Directly add the antiserum directly to the surface of the protein bands after electrophoresis, and the antigen reacts directly with the corresponding antibody to form a precipitate. Unbound free antigens or antibodies are washed away, resulting in the appearance of a specific protein that has been bound and fixed.[5]

In clinical practice, the accurate identification of monoclonal protein (M-spikes) in electrophoresis patterns is highly correlated with the experience and interpretation time of laboratory technicians. However, there are variations in the interpretation standards for M-spikes among technicians with different levels of experience. Therefore, there is a need to develop an Artificial Intelligence (AI) system for the automated interpretation of SPE images.

For the application of AI in the field of image recognition, the most advanced approach currently is based on deep convolutional neural networks (CNN) for image analysis.[6] The network has been applied in various fields such as clinical histopathology image analysis, Magnetic Resonance Imaging, and X-ray imaging,[7,8] providing high recognition accuracy and great convenience for clinical disease diagnosis. We propose an AI system based on a multilayer neural network for automatic recognition of serum protein electrophoresis (AIRSPE). It takes SPE images as input and predicts the presence of M-spike-related patterns, achieving expert-level performance within 15 ms. In this study, we selected MobileNetv2 as the deep learning CNN model, suitable for the interpretation of SPE images. MobileNetv2 is a lightweight CNN proposed by

Received: October 12, 2024
Revised: January 19, 2025
Accepted: January 27, 2025
Published: February 3, 2025

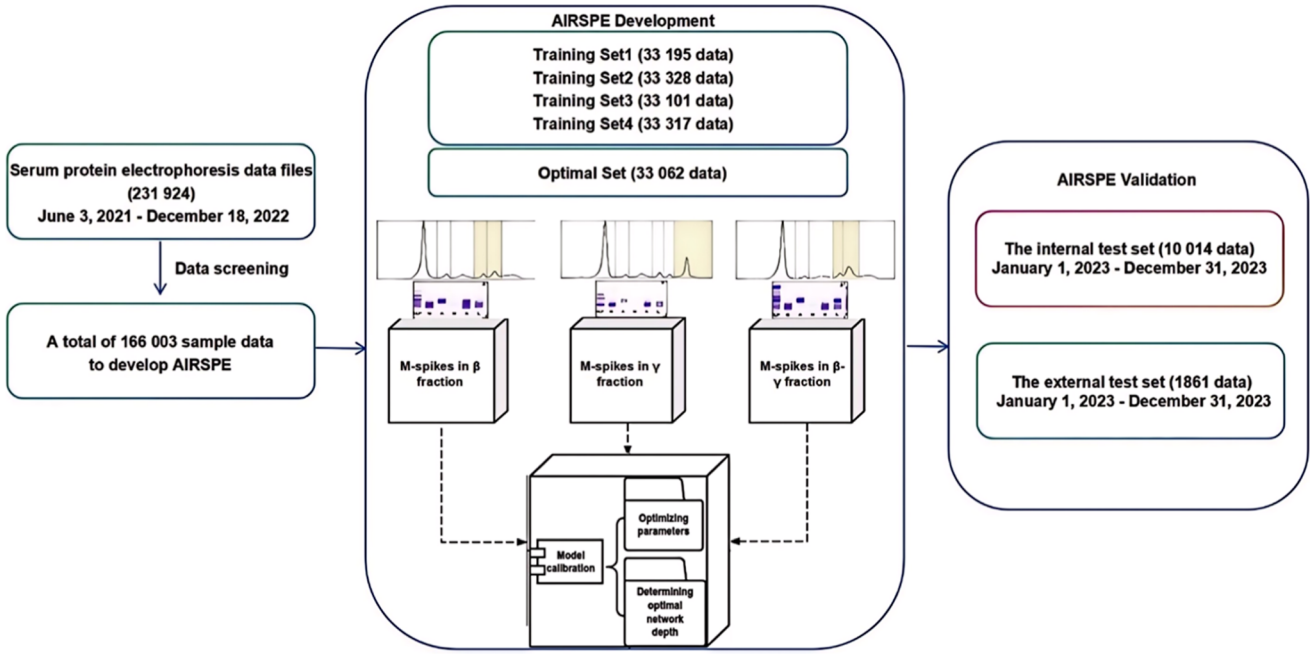


Figure 1. Framework for data processing, AIRSPE development and validation.

the Google team in 2018.[9] Compared with the MobileNetv1 version, the model is smaller (about 20% reduction in parameters) and has better accuracy. Additionally, MobileNetv2 introduces a different residual structure from ResNet, where the convolution is first expanded and then reduced, known as an inverted residual block. The main features of MobileNetv2 include: 1 the use of depthwise separable convolution instead of regular convolution to reduce the computational and parameter complexity of the model; 2 the introduction of inverted residuals to address the problem of gradient vanishing; 3 the use of linear functions (Hardtanh, Relu6) instead of the Relu function to reduce the loss of low-dimensional feature information.

The reported assays about expert-level interpretation of serum protein electrophoresis through deep learning used manual interpretation as the gold standard. Although senior electrophoresis technicians were selected for manual interpretation, they still had an error rate faced with high workload. So, our research used IFE as the gold standard in the internal test and external test groups to maximize accuracy.

2. MATERIALS AND METHODS

2.1. Cohort and Data. SPE testing data (166,003 cases, 57.24 ± 12.87 years old), which were screened from 231,924 data files from the Department of Laboratory Medicine Zhongshan Hospital, Fudan University, from June 3, 2021 to December 18, 2022 were randomly divided into 5 data sets: (a) 4 training sets (80%, with each set containing 33,195, 33,328, 33,101, and 33,317, sample cases, respectively) used to optimize model parameters and determine the optimal network depth; (b) 1 optimal set (20%, 33,062 sample cases) used to evaluate the performance of AIRSPE (Figure 1). The 166,003 electrophoresis data were initially obtained by a dual-person dual-check approach (2 technicians were requested to provide a categorical label (0 or 1) rather than free text for each of the electrophoresis images, and was blind to the others' labeling), and during the review period, a senior third-party conducted a reaudit and confirmation of the data (if there were different opinions, the final decision was made based on the consensus of the majority). The negative SPE data were confirmed by hematologists from Zhongshan Hospital, ruling out hematologic disorders potentially involved in the clinical diagnosis. Another 10,014 cases of SPE, confirmed by IFE, from January 1, 2023 to December 31, 2023 were selected as the internal test set (Figure 1).

Besides, 1861 samples (54.83 ± 12.44 years old) from the Department of Laboratory Medicine, Xiamen Hospital, Fudan University, from January 1, 2023 to December 31, 2023, were chosen as the external test set. The eligibility criteria were: (1) the patient had an IFE result; (2) the data did not have the same barcode.

This study was approved by the Ethics Committee of Zhongshan Hospital Affiliated to Fudan University (B2022-438), exempting patients from informed consent.

2.2. Measurements. The Sebia Capillary3 and Sebia Hydrasys 2 agarose gel electrophoresis systems were used for the SPE and IFE, respectively. Quality control materials, including normal and abnormal capillary electrophoresis components, were purchased from Sebia, France. Daily internal quality control was performed after calibration for the sample testing.

2.3. Data Preprocessing. Data preprocessing is a crucial step in AI applications, aiming to ensure the quality and consistency of raw data to provide a reliable foundation for model training and modeling. The data preprocessing workflow for SPE data includes operations such as image denoising, image enhancement, and image normalization in order to facilitate subsequent modeling and training.

In our study, data files with a length of zero, as well as data with insufficient length or invalid format, were excluded. In addition, data containing information about nonpatient specimens, such as quality control data, were also excluded. To ensure data consistency, we decoded the data in chronological order, and in cases where the same barcode appeared multiple times, only the last data were retained as valid data to eliminate interference from unstable scanning

https://doi.org/10.1021/acsomega.4c09327
ACS Omega 2025, 10, 5770−5777

data of the instrument. For the specimens collected from the same patient at different times, we also used the last data.

2.4. Deep Learning Convolutional Neural network (CNN) Model. To determine the optimal model for typical 5-fold cross-optimization training in order to maximize the suppression of noisy data and obtain optimal parameters, various deep learning architectures were compared, including Mobilenet, Resnet, and VGG.[10] After comparing their interpretive effectiveness, the most suitable CNN model were selected. Object-level metrics, such as F1-score (F1), precision (PREC), recall (REC), loss value (LOSS), accuracy (ACC), and interpretation time (IT) were used to evaluate the stability and generalization ability of these models. The synthesis of these parameters is closely linked to the overall model performance.

2.5. Training and Optimal Sets. Following the steps of CNN model establishment, the internal parameters of each layer of MobileNetV2 were configured. We generated 5-fold cross-training sets, which were used to complete the learning of AIRSPE, optimize model parameters, and determine the optimal network depth. Four sets were used as training sets, while one set was used as the optimal set. When displaying or training, we converted each sample into a high-definition image file with a size of 300 × 300 pixels. Based on deep learning methods, we established a CNN recognition model and trained the model using the labeled image files for classification (Figure 1). The final output of the system follows the principle of the majority rule for summarization and arbitration. After training, the optimal set was used to determine the optimal thresholds for AIRSPE.

2.6. Evaluations in Internal and External Test Sets. A sample data set of 10,014 SPE results were collected as an internal test set to validate the performance of AIRSPE. The internal test set required all to have IFE results, and the final IFE result was used as the criterion for judging the presence of M-spikes.

Manual interpretation of the internal test set was conducted by five experienced electrophoresis technicians with different years of experience (each with more than 5 years' experience; each technician was requested to provide a categorical label (0 or 1) rather than free text for each of the electrophoresis images, and was blind to the others' labeling; the final decision was made based on the consensus of the majority if there were different opinions). The five technicians conducted a manual blind review of the internal test set data (the judgment time required was within 30 s).

According to the internal test set, the performance of AIRSPE in identifying M-spikes was evaluated using sensitivity (TPR), specificity (TNR), false positive rate (FPR), false negative rate (FNR), positive predictive value (PPV), and negative predictive value (NPV), which were used to compare with the results of manual interpretation. In comparison, performance metrics were also evaluated on the external test set that included all of the samples run through another laboratory.

Number of samples, depicted as percentages that were detected by AIRSPE and manual interpretation at different M-spike concentrations, were compared in the positive data validated by IFE. Furthermore, the M-spikes detected by AIRSPE but not manual interpretation were analyzed based on different protein fractions (β, β+γ, and γ).

To reflect the improvement of the AIRSPE method in M-spike identification, the positive and negative numbers of M-spikes in different fractions (β, β+γ, and γ) of the internal test set were counted. The Receiver Operating Characteristic (ROC) curves were used to compare the area under the curve (AUC) of the AIRSPE method with the manual interpretation, and the net reclassification improvement (NRI) of each group was calculated to evaluate the potential of the AIRSPE method as a substitute for the manual interpretation.

2.7. Concordance of AIRSPE, Manual Interpretation, and IFE. The performance of AIRSPE and manual interpretation in screening M-spikes was verified by the concordance of the AIRSPE between IFE, manual interpretation between IFE, and AIRSPE between manual interpretation.

2.8. Statistical Analysis. Image analysis was performed using an NVIDIA GeForce GTX 2080 Ti (11GB) GPU, with each image taking 15 ms. The activation function used is the Hardtanh function, $f(x) = \max(-1, \min(1, x))$. Data analysis was conducted by using Python 3.6.9 and SPSS 27.0. Chi-square test was used for intergroup comparison of categorical data. The independent sample t-test was used for intergroup comparison of normally distributed continuous data, and the nonparametric rank sum test was used for intergroup comparison of non-normally distributed continuous data. The ROC curves were used to compare the ability of the AIRSPE method and manual method in identifying M-spikes. $P < 0.05$ or $P < 0.01$ indicates statistical significance.

3. RESULTS AND DISCUSSION

3.1. Data Preprocessing. All SPE instrument communication decoding data files (231,924) from June 3, 2021 to December 18, 2021 were selected as research objects. After data preprocessing and verification by senior third-party technicians and hematologists, 166,003 sample data were used in our experiment. The data were divided proportionally to form 5-fold cross-training and optimal sets for optimizing the model. For training sets, if there were more negative data than positive data, then all positive data were included, while only a batch of negative data was selected to participate in the training to balance the data.

3.2. Deep Learning Convolutional Neural Network Model. The adoption of CNN as the backbone network is driven by the necessity to efficiently process a vast amount of image data sets. The processing stage includes establishing multiscale feature maps, employing algorithms and selection strategies for generating and identifying target candidate boxes to accurately select high-quality target candidate areas, and utilizing a classification network to categorize these target areas while concurrently calculating their pixel information.

On the representative test set, the MobileNetv2 performed with high F1-score of 84.60%, precision of 76.20%, recall of 95.20%, loss of 26.80%, accuracy of 89.48%, and less interpretation time of 14 ms, which is most suitable for interpreting our CE images (Table 1).

Table 1. Performance of Different Models

	MobilenetV2	Resnet18	VGG19
F1-score (%)	84.60	84.30	81.70
precision (%)	76.20	75.90	73.30
recall (%)	95.20	94.80	91.70
loss (%)	26.80	27.00	31.20
accuracy (%)	89.48	87.70	81.20
interpretation time (ms)	14.00	29.20	26.40

https://doi.org/10.1021/acsomega.4c09327
ACS Omega 2025, 10, 5770−5777

In addition, the deep learning software used was developed in-house in order to meet the data analysis capacity within the laboratory. The adoption of MobileNetv2 improves the analytical efficiency without compromising the accuracy of the analysis of SPE.

3.3. Training and Optimal Sets. The internal parameters are listed in Table S1. Conv3 × 3 and Conv1 × 1 represent the convolutional layers, while the Bottleneck Layer is composed of inverted residual blocks.

The ratio of training sets (132,941) to the optimal set (33,062) is 4:1 (Table S2). There is no significant difference in the loss values of the models in the training set and optimal set ($P > 0.05$) (Figure S1). The average recall rate of each set of models is 95.20%.

According to the existing rules for manual interpretation of M-spikes: (1) abnormal peak patterns appear in the γ, β+γ, or β fractions; (2) the percentage of concentration in the β1 fraction is >9.5%; (3) the percentage of concentration in the β2 fraction is >12.1%; the convolutional layer identifies different features of the images through different convolutional kernels. After feature extraction and steps such as the Hardtanh activation function of the neural network, deep learning is performed. The optimal parameter configuration and activation function k value are obtained while reducing the residual of the neural network. A total of 33,062 cases in the optimal set were analyzed, with a positivity rate of 4.30%, which was not significantly different from the positivity rate in the training set (4.39%) ($P > 0.05$) (Table S2).

3.4. Internal and External Test Sets. A total of 10,014 cases from the internal testing sets were confirmed by IFE. The accuracy of AIRSPE was found to be 89.48% (8961/10,014), while the accuracy of manual interpretation was 88.05% (8817/10,014). The sensitivities of the AIRSPE and manual interpretation were 95.21% (2905/3051) and 68.76% (2098/3051) ($P < 0.05$), respectively, while the specificities were 86.97% (6056/6963) and 96.52% (6721/6963), respectively (Figure 2A).

As TPR of AIRSPE was higher than that of manual interpretation, the positive data validated by IFE (3051) were further analyzed according to different concentrations of M-spikes (g/L) (<0.5, [0.5,1), [1,2), [2,5), [5,10) and ≥10). The TPR rose with the increase of M-spike concentration. The data of M-spikes that were detected by AIRSPE but not by manual interpretation were predominantly found in the area of low-concentration M-spikes (especially when M-spikes were lower than 0.5 g/L; Figure 2B). To further analyze differences between AIRSPE and manual interpretation, the aforementioned data (807) were analyzed across different fractions (β, β+γ, and γ). The data most frequently missed by manual interpretation were concentrated in the γ-fraction, with M-spikes lower than 0.5 g/L (Figure 2C). At the same time, typical SPE spectra where AIRSPE interpreted M-spikes as positive but manual interpretation as negative were listed (Figure 3). Among them, the spectra of low-concentration M proteins in γ bands were the most common (Figure 3C).

The external test set-derived performance metrics yielded outcomes (including TPR, TNR, FPR, and NPV) that were equivalent to those derived from the internal test set ($P > 0.05$) (Table 2), thereby assuaging concerns regarding the AIRSPE's generalizability and the potential risk of overfitting.

The AUC for AIRSPE was 0.911 (95% CI, 0.9055−0.9164), while for manual interpretation it was 0.826 (95% CI, 0.8179−0.8349), with a significant difference ($P < 0.05$) (Figure S2).

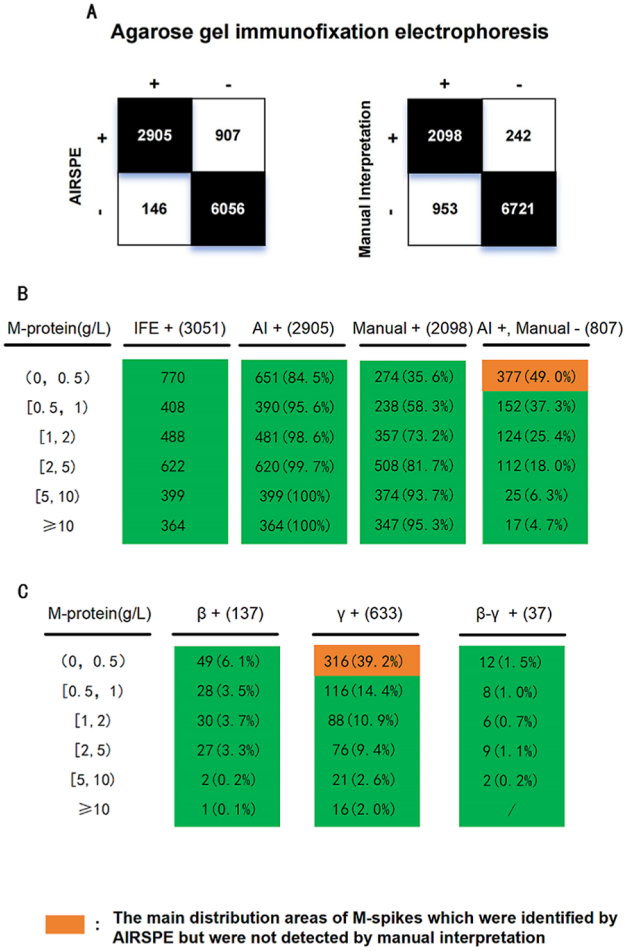


Figure 2. Comparison of the detection rate of monoclonal gammopathy by AIRSPE and manual interpretation in internal test sets (+: positive, −: negative). (A) Situation of positive and negative counts for AIRSPE, manual interpretation, and IFE. (B) In the case of a monoclonal protein at different concentration gradients, the number of positive detections by AIRSPE, manual interpretation, and the proportion of monoclonal protein in the positive IFE results. (C) In the case of a monoclonal protein at different concentration gradients, the distribution of results where AIRSPE reads as positive and manual interpretation reads as negative across different SPE fractions.

To further evaluate the performance of AIRSPE, ROC curves were plotted for different fractions of M-spikes. The AUC values of AIRSPE in β, β+γ, and γ bands were 0.845 (95% CI, 0.8249−0.8653), 0.900 (95% CI, 0.8802−0.9194), and 0.921 (95% CI, 0.9157−0.9259), respectively ($P < 0.05$). The AUC values of the manual interpretation method in the β, β+γ, and γ bands were 0.704 (95% CI, 0.6780−0.7294), 0.839 (95% CI, 0.8054−0.8735), and 0.843 (95% CI, 0.8339−0.8520), respectively ($P < 0.05$) (Figure S3). The AIRSPE method outperformed the manual method in the interpretation of M-spikes in all fractions.

After confirmation by the IFE, 3051 positive cases of internal test sets were used to calculate the NRI of the manual interpretation method and AIRSPE method. In different fractions, the NRI (%) calculated by the AIRSPE model was significantly higher than that of manual interpretation ($P < 0.01$), which indicated AIRSPE can identify M-spikes in SPE more accurately (Table 3).

3.5. Concordance of AIRSPE, Manual Interpretation, and IFE. The concordance of AIRSPE and IFE showed

https://doi.org/10.1021/acsomega.4c09327
ACS Omega 2025, 10, 5770−5777

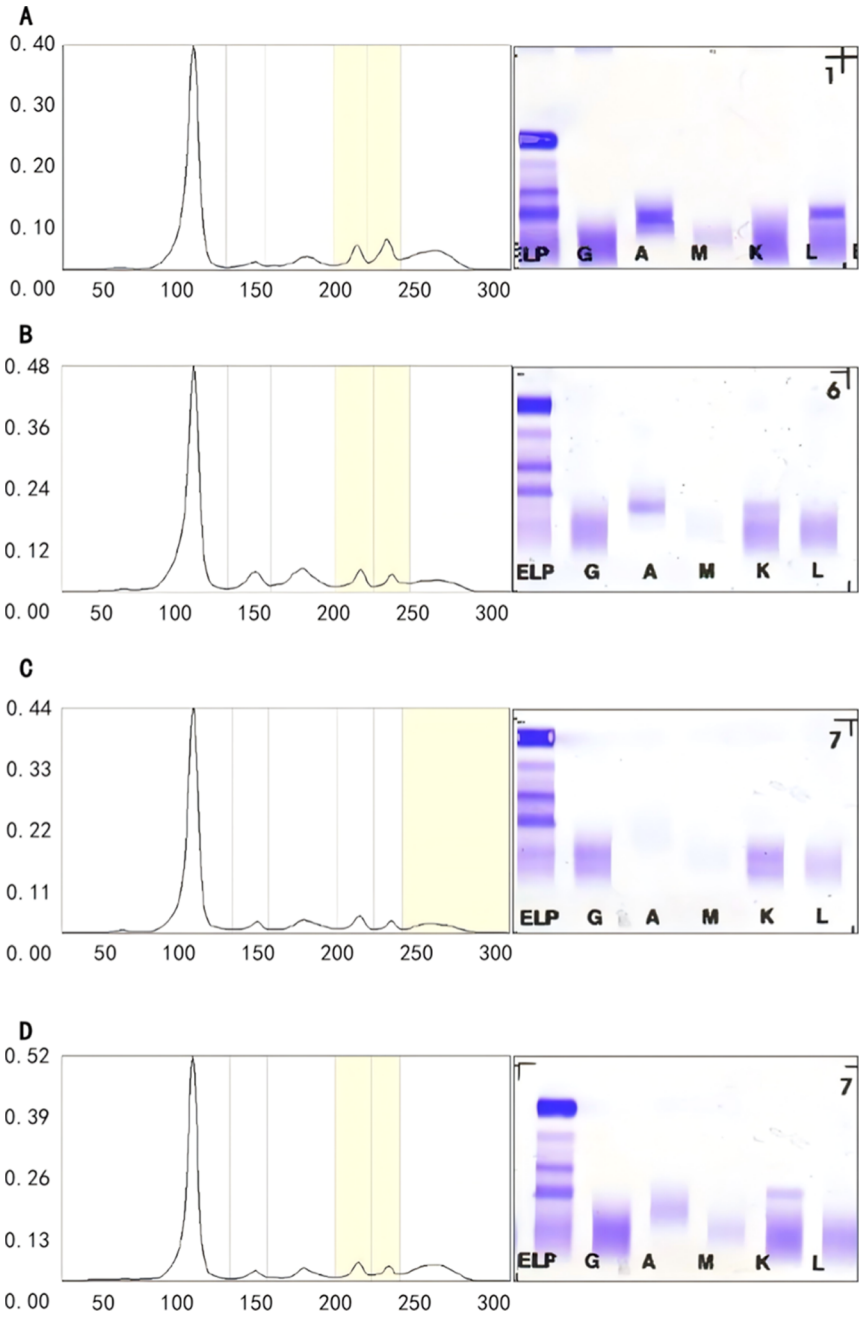


Figure 3. In the context of SPE, the electrophoresis patterns from AIRSPE that are interpreted as positive and the corresponding IFE patterns that are interpreted as negative. (A) In the β fraction, the monoclonal protein is interpreted as positive by AIRSPE and negative by manual interpretation, with IFE confirming it as positive. (B) In the $\beta-\gamma$ fraction, the monoclonal protein is interpreted as positive by AIRSPE and negative by manual interpretation, with IFE confirming it as positive. (C) In the γ fraction, the monoclonal protein is interpreted as positive by AIRSPE and negative by manual interpretation, with IFE confirming it as positive. (D) In the β fraction, the monoclonal protein is interpreted as positive by AIRSPE and negative by manual interpretation, with IFE confirming it as kappa light chain type.

Cohen's Kappa of 0.832, implying almost perfect agreement. Concordance for IFE and for the median results (according to the principle of minority obedience to majority) chosen by manual interpretation showed a Cohen's kappa of 0.699 ($P < 0.01$), indicating a substantial concordance between the technicians and IFE results.

Besides, concordance between manual interpretation and AIRSPE displayed a Cohen's kappa of 0.562, demonstrating a moderate concordance between technicians and AIRSPE (Table S3).

3.6. Discussion. M-spikes are often associated with diseases such as monoclonal gammopathy and multiple myeloma.[11] Currently, CZE is the most commonly used screening method for M-spikes in clinical practice,[12] as it requires a small amount of serum specimen and provides fast results. In the identification of a large number of capillary electrophoresis patterns, humans are prone to visual fatigue, especially when the concentration of the M-spike is low and the peak shape changes are not obvious. In such cases, the introduction of AIRSPE significantly improves the effectiveness.

https://doi.org/10.1021/acsomega.4c09327
ACS Omega 2025, 10, 5770−5777

Table 2. Evaluations of AIRSPE and Manual Interpretation on Internal and External Test Sets

	IFE+	IFE−	true positive rate (%)	true negative rate (%)	false positive rate (%)	false negative rate (%)	negative predictive value (%)	positive predictive value (%)
				Internal Test Set (10,014 Data)				
AIRSPE+	2905	907	95.21[a]	86.97[a]	13.03[a]	4.79	97.65[a]	76.21
AIRSPE-	146	6056						
Manual interpretation+	2096	242	68.76	96.52	3.48	31.30	87.56	89.65
Manual interpretation-	955	6721						
				External Test Set (1861 Data)				
AIRSPE+	318	193	89.58[a]	87.18[a]	12.82[a]	10.42	97.26[a]	62.23
AIRSPE-	37	1313						
Manual interpretation+	286	155	80.56	89.71	10.29	19.44	95.14	64.85
Manual interpretation-	69	1351						

[a]The performance metrics of AIRSPE in the internal test set and external test set: $P > 0.05$.

Table 3. Comparison of Net Reclassification Improvement Values of Manually Interpreted and AIRSPE Interpreted M-Spikes in Different Fractions

zones	model	NRI (%)	95% CI (%)	P value
β fraction	AIRSPE	37.85	32.79−42.90	<0.01
	Manual interpretation	−9.57	−10.30∼−8.84	
β+γ fraction	AIRSPE	21.64	15.47−27.81	<0.01
	Manual interpretation	−9.57	−10.30∼−8.84	
γ fraction	AIRSPE	25.12	23.43−26.81	<0.01
	manual interpretation	−9.55	−10.28∼−8.82	

The successful application of deep learning relies on the availability of large-scale data sets that are suitable for training and optimization. The samples used for the deep learning were carefully selected, excluding nonpatient specimens such as quality control data and duplicate barcode specimens. The training sets were further refined by removing data that conflicted with the results of AIRSPE and IFE, resulting in a total of 166,003 valid samples. The training results were comparable to those of human experts with years of experience. The large amount of data was sufficient for training and optimization, reducing the error rate in AIRSPE's interpretation of M-spikes. Additionally, appropriate parameter settings during modeling helped to reduce the loss value during convolution, which were enhanced by multiple training sets.

The MobileNetv2 was selected for deep learning and analysis of capillary electrophoresis profiles. Due to the moderate amount of data in this study, low-volume and efficient models such as MobileNet are suitable for these data sets. The depth-wise separable convolution accurately extracts the boundary contours of the electrophoresis patterns, enabling multiscale region of interest (RoI) recognition. The pooling layer compresses and classifies the images, extracting image features, and outputting recognized images. This has the potential to enhance the diagnostic value of capillary electrophoresis in M-spikes related diseases.

A short time ago, Floris Chabru et al. created a model named "SPECTR" for the interpretation and analyses of SPEs.[13] The interpretation of SPECTR was compared to that of 9 independent experts from Biochemistry laboratories of 7 hospital centers on 50 different samples, which avoided "the black box effect". What sets Chabru's research apart is that the evaluation of AIRSPE comparing results of the internal test set with matched IFE results. The AIRSPE demonstrated better sensitivity and NRI. As the gold standard for M-spike assessment, the internal test set with IFE results enhanced the reliability and clinical utility of AIRSPE.

As a screening method for M-spike-related diseases, the AIRSPE established in this study focuses on the sensitivity of M-spike detection. The sensitivity of AIRSPE (95.21%) is higher than manual interpretation (68.76%), which is consistent with our intention to ensure that AIRSPE does not miss any suspicious M-spike. Comparison with the diagnostic performance of the manual method reveals that the internal test set confirmed by AIRSPE has a good AUC (0.911) and accuracy rate (89.48%), indicating that it may be an ideal M-spike prediction model. AIRSPE and manual interpretation show the most significant differences in the low-concentration γ fraction of SPE. We conducted an in-depth study of the areas where AIRSPE excels and found that when the concentration of M-spikes is lower than 0.5 g/L, it is difficult for the naked eye to discern changes in the shape of the spectrum, while AIRSPE, by convoluting and analyzing the acquired spectrum, can identify minor changes in the peak shape.

According to the different fractions of M-spikes (β, β+γ, and γ fractions), AIRSPE shows good AUC values in all fractions (0.845, 0.900, and 0.921), with the highest diagnostic performance in the γ fraction. Abnormal proliferation of monoclonal immunoglobulins in the γ fraction is mainly IgG type (one of the category of Immunoglobulin (Ig)), accounting for 75% to 80% of the total serum immunoglobulins.[14] In monoclonal-related diseases, high concentrations of IgG are often found, and the electrophoretic pattern in the γ fraction usually shows narrow, high, and sharp protein peaks, which is obvious for manual and AIRSPE interpretation. The β fraction has a lower incidence rate of M-spikes, and the abnormal proliferation of monoclonal immunoglobulins is mainly IgA type, accounting for 10% to 20% of the total serum immunoglobulins. The shape change in the β fraction on the electrophoretic pattern is not significant, resulting in a lower AUC compared with the γ fraction. IgA monoclonal proteins usually migrate to the β fraction,[15] and strong polyclonal increases often lead to the fusion of the β2 and γ fractions, referred to as β−γ bridge or β−γ fusion,[16] often associated

https://doi.org/10.1021/acsomega.4c09327
ACS Omega 2025, 10, 5770−5777

with alcoholism, liver cirrhosis, respiratory tract infections, or rheumatoid arthritis. Monoclonal immunoglobulins in the $\beta-\gamma$ fraction are usually accompanied by polyclonal IgA,[17] which is difficult to accurately identify using manual interpretation. AIRSPE has a better diagnostic performance and is suitable for the determination of M-spikes in this fraction.

NRI was used to quantitatively evaluate the net improvement of AIRSPE in diagnosing M-spike-related diseases. AIRSPE showed higher NRI in each M-spike fraction, indicating that AIRSPE can accurately identify M-spikes in different fractions of SPE, and has the potential for clinical routine interpretation of M-spikes. The MobileNetv2 extracts key information from SPE's patterns through multiple convolutional layers and then converts the images into a three-dimensional tensor, performs precise calculations on the target regions, and finally converts them into binary analysis to interpret the results. In comparison, manual interpretation relies on the examiner's visual observation of the electrophoretic patterns and analysis based on past experience, which often makes it difficult to accurately judge low concentrations of M-spikes.

One limitation of AIRSPE is that the diagnostic performance in the β fraction is not significantly improved. To address this issue, further research can be conducted to reanalyze the identification of positive samples in the SPE of the β fraction, combined with the results of IFE and immunotyping tests, in order to optimize the parameters of the MobileNetv2 and improve the detection rate of M-spikes. In addition, the 166,003 data included in training sets were reviewed by a team of two physicians using dual-reading and posthoc review by a senior physician from a third party, but no corresponding data points were validated using IFE, which may result in a certain degree of sensitivity loss.

It is necessary to explore the identification method of light chain type M-spikes in SPE. The light chain type accounts for approximately 18% of monoclonal gammopathy of undetermined significance (MGUS),[18] and its proportion in clinical diagnosis has been increasing year by year. However, it is often difficult to detect in the SPE spectrum. It is expected that AIRSPE will overcome this challenge in its future development.

4. CONCLUSIONS

The establishment of AIRSPE is a lengthy process that involves selecting a neural network model suitable for the interpretation of SPE images. After training with large amounts of data, optimization, internal and external test sets, small-scale experiments can be conducted on clinical patient samples. Compared to the manual judgment of M-spikes, AIRSPE has better interpretation efficiency mainly in the γ-fraction with low-concentration M-spikes (<0.5 g/L), indicating its potential for assisting in the clinical screening of M-spike diseases.

■ ASSOCIATED CONTENT

Data Availability Statement

The data and materials used in this study can be obtained upon request from the authors. To request access to the data and materials, please contact the corresponding author. Once the request is reviewed and approved, the data and materials will be made available to you.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.4c09327.

Internal parameters of each layer of the MobileNet v2 model; number of negative and positive cases in training and validation sets; comparison of concordance between AIRSPE, manual interpretation, and IFE; variation of model loss in each group of training and optimal sets; ROC curves of AIRSPE versus manual interpretation for identifying monoclonal proteins; and ROC curves in different zones of AIRSPE versus manual interpretation for identifying monoclonal proteins; (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Wei Guo − *Department of Laboratory Medicine, Zhongshan Hospital, Fudan University, Shanghai 200032, China*; Email: guo.wei@zs-hospital.sh.cn

Beili Wang − *Department of Laboratory Medicine, Zhongshan Hospital, Fudan University, Shanghai 200032, China*; Email: wang.beili1@zs-hosptial.sh.cn

Wenqi Shao − *Department of Laboratory Medicine, Zhongshan Hospital, Fudan University, Shanghai 200032, China*; Email: shao.wenqi@zs-hosptial.sh.cn

Authors

Ruojian Shen − *Department of Laboratory Medicine, Zhongshan Hospital, Fudan University, Shanghai 200032, China*; orcid.org/0000-0002-9924-638X

Yuyi Hu − *Department of Laboratory Medicine, Zhongshan Hospital, Fudan University, Shanghai 200032, China*

Qun Wu − *Department of Laboratory Medicine, Zhongshan Hospital, Fudan University, Shanghai 200032, China*

Jing Zhu − *Department of Laboratory Medicine, Zhongshan Hospital, Fudan University, Shanghai 200032, China*

Wen Xu − *Department of Laboratory Medicine, Zhongshan Hospital, Fudan University, Shanghai 200032, China*

Baishen Pan − *Department of Laboratory Medicine, Zhongshan Hospital, Fudan University, Shanghai 200032, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.4c09327

Author Contributions

R.S., Y.H., and Q.W. contributed equally to this work and they are cofirst authors. W.G., B.W., and W.S. contributed equally to this work and they are cocorresponding authors. Beili Wang, Paishen Pan, and Wei Guo conceived and designed the experiments. Ruojian Shen, Yuyi Hu, and Wenqi Shao performed the research, conducted the data analyses, and wrote the manuscript. Jing Zhu, Wen Xu, and Qun Wu contributed to the data collection. Wenqi Shao, Ruojian Shen, and Yuyi Hu revised the manuscript. All authors read and approved the final manuscript.

Funding

The following are acknowledged: The National Natural Science Foundation of China (81972000, 82172348); The National Nature Foundation Youth Project (81902139); Specialized Fund for the clinical researches of Zhongshan Hospital affiliated Fudan University (020ZSLC54); the constructing project of clinical key disciplines in Shanghai (shslczdzk03302); Key Medical College of Baoshan District,

https://doi.org/10.1021/acsomega.4c09327
ACS Omega 2025, 10, 5770−5777

Shanghai (BSZK-2023-A18); Xiamen Medical and Health Key Project (YDZX20193502000002).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Ruojian Shen and W.S. for their assistance in collecting the data.

■ REFERENCES

(1) Keren, D. F.; Bocsi, G.; Billman, B. L.; Etzell, J.; Faix, J. D.; Kumar, S.; Lipe, B.; McCudden, C.; Montgomery, R.; Murray, D. L.; et al. Laboratory Detection and Initial Diagnosis of Monoclonal Gammopathies. *Arch. Pathol. Lab. Med.* **2022**, *146* (5), 575−590.

(2) Landgren, O.; Kyle, R. A.; Pfeiffer, R. M.; Katzmann, J. A.; Caporaso, N. E.; Hayes, R. B.; Dispenzieri, A.; Kumar, S.; Clark, R. J.; Baris, D.; et al. Monoclonal gammopathy of undetermined significance (MGUS) consistently precedes multiple myeloma: a prospective study. *Blood* **2009**, *113*, 5412−5417.

(3) Holler, F. J.; Skoog, D. A.; Crouch, S. R. *Capillary Electrophoresis. In: Principles of instrumental analysis*; Cengage Learning, 2007, p 794.

(4) Hage, D. S. An Overview of CE in Clinical Analysis. *Methods Mol. Biol.* **2019**, *1972*, 3−11.

(5) Ritchie, R. F.; Smith, R.; Immunofixation, I. General principles and application to agarose gel electrophoresis. *Clin. Chem.* **1976**, *22* (4), 497−499.

(6) Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521* (7553), 436−444.

(7) Lakhani, P. Deep Convolutional Neural Networks for Endotracheal Tube Position and X-ray Image Classification: Challenges and Opportunities. *J. Digital Imaging* **2017**, *30*, 460−468.

(8) Rajkomar, A.; Lingam, S.; Taylor, A. G.; Blum, M.; Mongan, J. High-Throughput Classification of Radiographs Using Deep Convolutional Neural Networks. *J. Digital Imaging* **2017**, *30* (1), 95−101.

(9) Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L. *Inverted Residuals and Linear Bottlenecks: Mobile Networks for Classification, Detection and Segmentation*. 2018.

(10) Dalvi, P. P.; Edla, D. R.; Purushothama, B. R. Diagnosis of Coronavirus Disease From Chest X-Ray Images Using DenseNet-169 Architecture. *SN Comput. Sci.* **2023**, *4* (3), 214.

(11) Cao, F.; Zhang, R.; Xu, L.; Liu, M.; Yuan, Y. Application of Capillary Electrophoresis in Monoclonal Gammopathies and the Cutoff Value of Monoclonal Protein in Differential Diagnosis of Multiple Myeloma and Other Monoclonal Gammopathies. *Ann. Clin. Lab. Sci.* **2021**, *51* (3), 400−407.

(12) Harstad, R. K.; Johnson, A. C.; Weisenberger, M. M.; Bowser, M. T. Capillary Electrophoresis. *Anal. Chem.* **2016**, *88* (1), 299−319.

(13) Chabrun, F.; Dieu, X.; Ferre, M.; Gaillard, O.; Mery, A.; Chao de la Barca, J. M.; Taisne, A.; Urbanski, G.; Reynier, P.; Mirebeau-Prunier, D. Achieving Expert-Level Interpretation of Serum Protein Electrophoresis through Deep Learning Driven by Human Reasoning. *Clin. Chem.* **2021**, *67* (10), 1406−1414.

(14) Yasunaga, Y.; Bonilla-Palacios, J. J.; Shinomura, Y.; Kanayama, S.; Miyazaki, Y.; Matsuzawa, Y. High prevalence of serum immunoglobulin G antibody to Helicobacter pylori and raised serum gastrin and pepsinogen levels in enlarged fold gastritis. *Can. J. Gastroenterol.* **1997**, *11* (5), 433−436.

(15) Katzmann, J. A.; Willrich, M. A.; Kohlhagen, M. C.; Kyle, R. A.; Murray, D. L.; Snyder, M. R.; Rajkumar, S. V.; et al. Monitoring IgA multiple myeloma: immunoglobulin heavy/light chain assays. *Clin. Chem.* **2015**, *61* (2), 360−367.

(16) Chan, P. C.; Lem-Ragosnig, B.; Chen, J. Diagnostic implications of enumerating and reporting beta fraction(s) for the detection of beta-migrating monoclonal immunoglobulins in serum protein electrophoresis. *Clin. Biochem.* **2018**, *53*, 77−80.

(17) Omar, N.; Madwani, K.; Moideen, P.; Manthei, D. M.; Keren, D. F.; Singh, G. Accurate Quantification of Monoclonal Immunoglobulins Migrating in the Beta Region on Protein Electrophoresis. *Lab. Med.* **2022**, *53* (2), 138−144.

(18) Singh, G.; Xu, H. Light Chain Predominant Intact Immunoglobulin Monoclonal Gammopathy Disorders: Shorter Survival in Light Chain Predominant Multiple Myelomas. *Lab. Med.* **2021**, *52* (4), 390−398.

https://doi.org/10.1021/acsomega.4c09327
ACS Omega 2025, 10, 5770−5777