*IOSR Journal of Computer Engineering (IOSR-JCE)*
*e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 28, Issue 5, Ser. I (Sep. – Oct. 2026), PP 26-31*
*www.iosrjournals.org*

# Lightweight CNN–Transformer and Explainable AI for Real-Time Counterfeit Currency Detection and Classification

Kumud Wasnik[1], Dr. Ankit Temurnikar[2], Dr. Latha Gadepaka[3]
*[1]Research Scholar, Department of CSE, Madhyanchal Professional University, Bhopal, MP (India)*
*[2]Supervisor, Department of CSE, MPU Bhopal.MP (India)*
*[3]Co-Supervisor, Department of CSE-Cyber Security, School of Engineering, Malla Reddy University, Hyderabad, Telangana State*

***Abstract***
*This paper proposes a deployment-oriented artificial intelligence framework for real-time counterfeit currency authentication and denomination classification using hyperspectral imaging. Building on spectral-spatial representations, the proposed system combines informative spectral-band selection, a lightweight three-dimensional convolutional stem, spectral-spatial attention, and a compact transformer encoder. The architecture is formulated as a multi-task network with authentication and denomination heads and an optional suspicious-region localization branch. Explainable AI is incorporated through Grad-CAM and attention-based visualizations so that the system can indicate regions contributing to a decision. To support practical deployment, model compression, quantization, knowledge distillation, and inference profiling are incorporated into the experimental design. The evaluation protocol measures not only classification quality but also latency, throughput, memory, parameter count, and model size. No experimental results are fabricated in this manuscript; the reported-result tables are designed to be completed from controlled experiments. The contribution is a reproducible blueprint for an edge-deployable HSI currency inspection system that jointly addresses authentication, classification, interpretability, and real-time constraints.*
***Keywords***
*Hyperspectral imaging; counterfeit currency; CNN-transformer; vision transformer; multi-task learning; explainable AI; edge AI; real-time classification; spectral-spatial attention.*

## I. Introduction

A practical counterfeit currency detector must satisfy two competing requirements: it must be accurate enough to distinguish genuine and counterfeit notes, and it must be fast enough for operational use. Hyperspectral imaging increases the information available to an AI system but also increases computational load. A model that achieves high accuracy but requires excessive memory or latency may not be suitable for automated cash handling.

This paper addresses this deployment gap through a lightweight hybrid CNN-transformer architecture. CNN layers provide efficient local spectral-spatial feature extraction, while a compact transformer models relationships among a reduced set of tokens. The system is further formulated as a multi-task learner that performs authentication and denomination classification simultaneously.

## II. Research Objectives

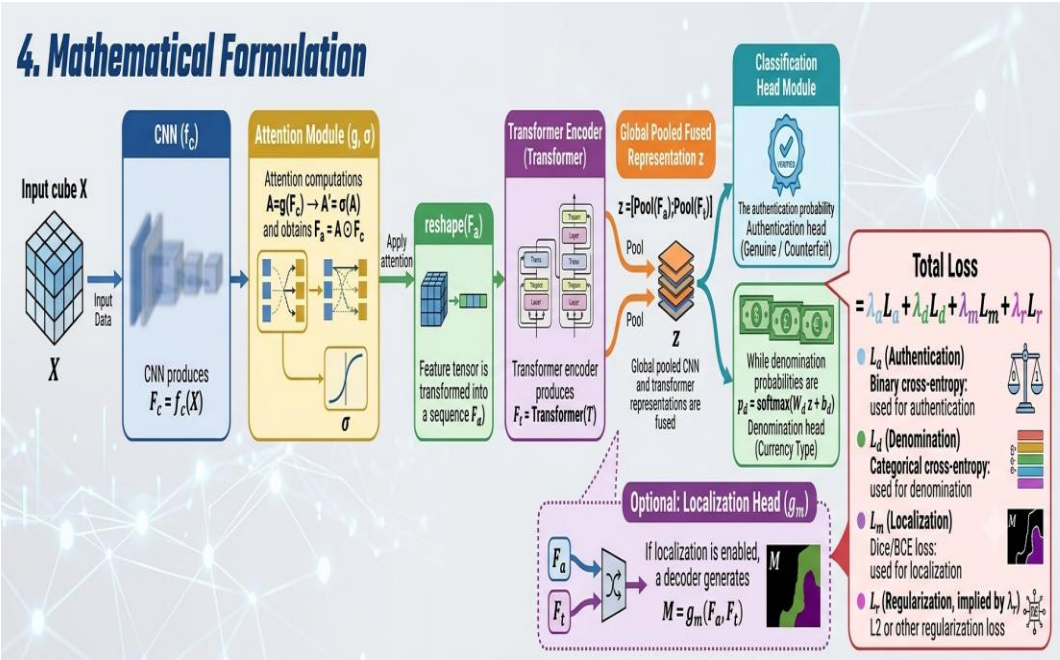
- Develop a compact HSI representation suitable for real-time inference.
- Combine local CNN features with global transformer relationships.
- Perform genuine/counterfeit authentication and denomination classification jointly.
- Provide optional suspicious-region localization and explainability.
- Quantify the trade-off between accuracy and computational cost.
- Evaluate quantization and compression for edge deployment.

## III. Proposed Architecture

The architecture consists of six stages:
(1) HSI preprocessing,
(2) informative-band selection,
(3) 3D-CNN feature extraction,
(4) spectral-spatial attention,

(5) compact transformer encoding, and
(6) multi-task prediction.


## IV. Mathematical Formulation




## V. Spectral-Spatial Attention

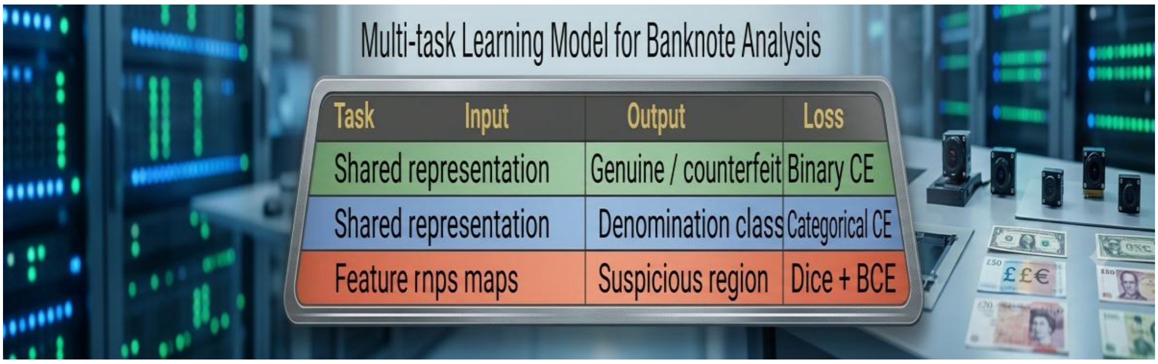
Spectral attention is used to suppress redundant bands and emphasize discriminative responses. Spatial attention can additionally focus on note regions that contain informative printing or material patterns. A dual attention module can be evaluated against spectral-only attention through an ablation study.

## VI. Transformer Tokenization

Instead of treating every hyperspectral pixel as a transformer token, which can become computationally expensive, the proposed system forms compact tokens from pooled spectral-spatial feature regions. This reduces the quadratic self-attention cost. Token dimension and token count should be selected through validation experiments, with real-time latency included as a selection criterion.

## VII. Multi-Task Learning

Authentication and denomination classification share the feature extractor but use separate prediction heads. This arrangement allows the model to learn features useful for both identifying a note and verifying its authenticity. A third localization head can be included when region annotations are available. The multi-task formulation should be compared against single-task authentication to determine whether auxiliary denomination learning improves generalization.




## VIII. Explainable AI

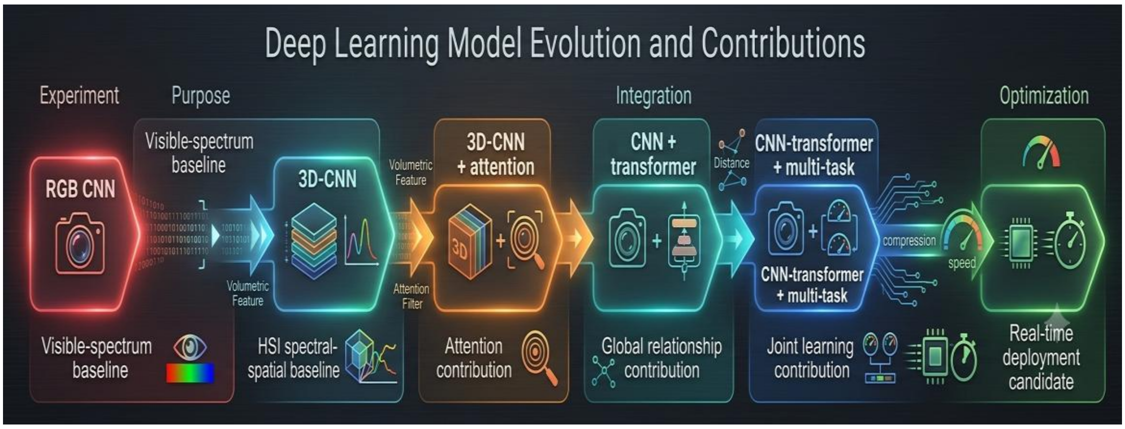
A high-stakes authentication model should provide evidence for its decision. Grad-CAM can highlight spatial regions associated with the predicted class. Transformer attention rollout can provide a complementary token-level interpretation. Spectral importance can be plotted across wavelengths to show which bands contributed strongly. Explanations should be treated as diagnostic evidence rather than proof of causality.

## IX. Real-Time Optimization

Real-time deployment requires optimization at several levels. First, band selection reduces input dimensionality. Second, depthwise or grouped operations can reduce convolutional cost. Third, token reduction limits transformer complexity. Fourth, post-training quantization or quantization-aware training can reduce memory and inference cost. Knowledge distillation can transfer performance from a larger teacher model to a smaller student model.

- Band reduction: retain only discriminative wavelengths.
- Token reduction: use compact pooled tokens.
- Pruning: remove low-contribution weights where supported.
- Quantization: evaluate FP32, FP16, and INT8 where hardware permits.
- Distillation: train a compact student using teacher predictions/features.
- Runtime optimization: profile preprocessing and inference separately.

## X. Experimental Protocol



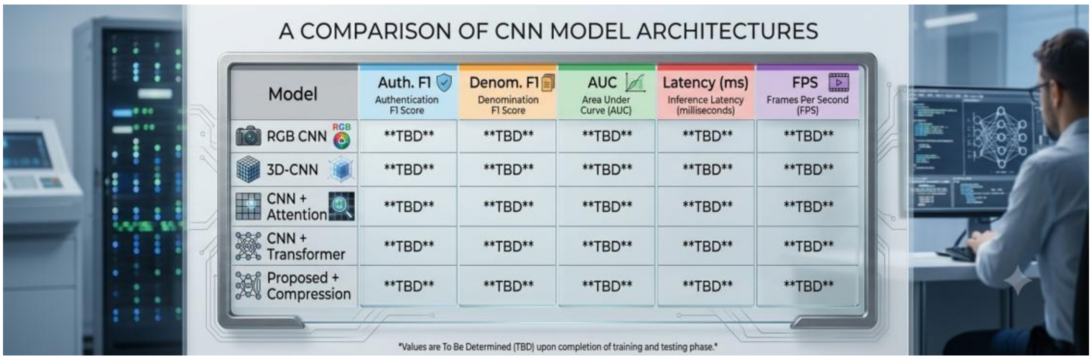

## XI. Dataset and Splitting

The dataset should contain multiple independent physical notes for each denomination and authenticity category. All captures from the same physical note must remain in one partition. A recommended initial split is 70% training, 15% validation, and 15% testing, although a repeated cross-validation design may be preferable for small datasets. The test set must remain untouched during architecture and hyperparameter selection.

## XII. Evaluation Metrics

Classification performance should include accuracy, precision, recall, specificity, macro-F1, ROC-AUC, and confusion matrices. Denomination performance should be reported per class and using macro-averaged metrics. Localization can be evaluated using IoU and Dice coefficient when ground-truth regions are available. Deployment performance should include end-to-end latency per note, model-only latency, throughput in notes per second, parameter count, FLOPs, RAM usage, model size, and power consumption when measurable.

## XIII. Results Tables



A COMPARISON OF CNN MODEL ARCHITECTURES

| Model | Auth. F1 Authentication F1 Score | Denom. F1 Denomination F1 Score | AUC Area Under Curve (AUC) | Latency (ms) Inference Latency (milliseconds) | FPS Frames Per Second (FPS) |
|---|---|---|---|---|---|
| RGB CNN | **TBD** | **TBD** | **TBD** | **TBD** | **TBD** |
| 3D-CNN | **TBD** | **TBD** | **TBD** | **TBD** | **TBD** |
| CNN + Attention | **TBD** | **TBD** | **TBD** | **TBD** | **TBD** |
| CNN + Transformer | **TBD** | **TBD** | **TBD** | **TBD** | **TBD** |
| Proposed + Compression | **TBD** | **TBD** | **TBD** | **TBD** | **TBD** |

*Values are To Be Determined (TBD) upon completion of training and testing phase.*

## XIV. Ablation Study


## XV. Error and Explainability Analysis

For each false positive and false negative, the study should compare the original HSI cube, selected-band images, attention maps, spectral importance curves, and note metadata. This analysis can reveal whether errors arise from worn genuine notes, unusual counterfeit printing, illumination changes, background artifacts, or sensor noise. The objective is to convert model failures into actionable improvements in data collection and architecture.

## XVI. Cross-Condition and Cross-Device Generalization

A deployment model should not be evaluated only under the exact conditions used during training. Testing should include held-out orientations and illumination conditions. If multiple sensors are available, a cross-device experiment should train on one device and test on another. Domain adaptation or normalization can then be studied if performance degrades.

## XVII. Security, Ethics, and Responsible Use

The system should be developed strictly for currency authentication and research. Counterfeit specimens must be handled under applicable legal and institutional controls. Published material should avoid details that could assist in reproducing security features. The system should be treated as an inspection aid and integrated with appropriate human or institutional verification processes for high-consequence decisions.

## XVIII. Discussion

The proposed architecture addresses a key tension in HSI AI: richer sensing increases discriminative information but also increases computation. Band selection and compact tokenization reduce the burden before transformer processing. Multi-task learning provides an additional supervisory signal through denomination recognition, while explainability supports model auditing. Nevertheless, real-time performance must be demonstrated on the intended deployment hardware rather than inferred from desktop training speed.

## XIX. Conclusion

This paper presents a lightweight CNN-transformer framework for real-time counterfeit currency detection and classification using hyper spectral imaging. The proposed design combines spectral-band selection, spectral-spatial attention, and compact transformer encoding, multi-task prediction, explain ability, and deployment optimization. The experimental framework explicitly measures both recognition quality and computational efficiency. Once evaluated on a sufficiently diverse physical-note dataset, the system can provide a strong research basis for edge-oriented currency authentication

## XX. Limitations and Future Work

- A larger and more diverse legally sourced dataset is required for strong generalization claims.
- Cross-sensor calibration and domain adaptation require further study.
- Real-world field evaluation should be conducted under approved operational conditions.
- Uncertainty estimation could identify low-confidence notes for manual inspection.
- Future work may investigate self-supervised HSI pretraining and federated/secure model updates.

## References

[1]. Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using deep belief networks," IEEE Transactions on Geoscience and Remote Sensing, 2014.
[2]. A.Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," ICLR, 2021.
[3]. J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," CVPR, 2018.
[4]. R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," ICCV, 2017.
[5]. A.Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," NeurIPS, 2012.
[6]. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," CVPR, 2016.
[7]. F. Chollet, "Xception: Deep learning with depthwise separable convolutions," CVPR, 2017.
[8]. Han, M., & Kim, J. (2019). Joint banknote recognition and counterfeit detection using explainable artificial intelligence. Sensors, 19(16), 3607. https://doi.org/10.3390/s19163607
[9]. Linkon, A. H. M., Labib, M. M., Bappy, F. H., Sarker, S., E-Jannat, M., & Islam, M. S. (2021). Deep learning approach combining lightweight CNN architecture with transfer learning: An automatic approach for the detection and recognition of Bangladeshi banknotes. arXiv preprint arXiv:2101.05081.
[10]. Ali, T., Salman, J., Ahmad, A., Mohammad, N., Muhammad, A., & Shoaib, S. M. (2019). DeepMoney: Counterfeit money detection using generative adversarial networks. PeerJ Computer Science, 5, e216.
[11]. Aseffa, D. T., Kalla, H., & Mishra, S. (2022). Ethiopian banknote recognition using convolutional neural network and its prototype development using embedded platform. Journal of Sensors, 2022, 1–18.
[12]. Brauwers, G., & Frasincar, F. (2023). A general survey on banknote recognition using deep learning. Expert Systems with Applications, 229, 120526.
[13]. Pruthvi, M. R., Saumya, Y. M., Kumari, B. J., Prajna, M., & Jithya, K. V. (2025). A deep learning approach for enhanced counterfeit currency detection and classification. In Proceedings of the International Conference on Mathematical Engineering, Modeling and Computer Science. Atlantis Press.
[14]. Alene, A. (2019). Designing Ethiopian banknote classification and counterfeit verification system: An optimal feature extraction and classification techniques. International Journal of Computer Applications, 178(20), 1–8.
[15]. Khan, M. A., et al. (2024). Jordanian banknote data recognition: A CNN-based approach with attention mechanism. Journal of King Saud University—Computer and Information Sciences, 36(4), 102038. https://doi.org/10.1016/j.jksuci.2024.102038
[16]. G., S., & Manoj, M. (2025). A CNN-driven multi-denomination Indian currency recognition and valuation system for real-time applications. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.6100679
[17]. Rao, N. A., & Loshma, G. (2025). Indian fake currency identification using CNN model. In Applications of Computational Intelligence in Management and Mathematics I. Springer.
[18]. Mohan, L., Janga, V., Chode, S. V., & Veeramani, V. (2026). Smart detection of Indian counterfeit currency notes using deep learning techniques. Security and Privacy, 9, e700.
[19]. Pavani, K., & Prasad, C. T. D. (2026). A robust framework for counterfeit currency detection using CNN. International Journal of Engineering Research and Science & Technology, 22(2), 3056–3063.
[20]. Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), 3–19.
[21]. Dosovitskiy, A., et al. (2021). An image is worth 16×16 words: Transformers for image recognition at scale. In International Conference on Learning Representations (ICLR).
[22]. Liu, Z., et al. (2021). Swin Transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 10012–10022.
[23]. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 4510–4520.
[24]. Howard, A. G., et al. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

[25]. Selvaraju, R. R., et al. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), 618–626.
[26]. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (NeurIPS), 4765–4774.