Figure Legend

Figure S1. A, Training curves of different frameworks of CNN model based on training dataset. Different frameworks revealed various speed of loss decline and tend to convergence. B, C, Validation curves of different frameworks of CNN model for recognition of atrophic gastritis and gastric intestinal metaplasia based on validation dataset.

Figure S1



 Table S1. Clinical characteristics of collected patients

|  |  |  |
| --- | --- | --- |
|   | Total Enrolled N=1,826 | Not Enrolled N=14,685 |
| Gender |  |  |
| Male  | 947 (51.9%) | 7416 (50.5%) |
| Female | 879 (48.1%) | 7269 (49.5%) |
| Age |  |  |
| >60 years | 461 (25.2%) | 3918 (26.7%) |
| 40-60 years | 975 (53.4%) | 7609 (51.8%) |
| <40 years | 390 (21.4%) | 3158 (21.5%) |
| Lesion |  |  |
| AG and/or GIM | 667 (36.6%) | 5,551 (37.8%)  |
| CNAG | 1,159 (63.6%)  | 9134 (62.2%) |

Table S2. Annotation of 529 images with AG and/or GIM by three expert endoscopists.

|  |  |
| --- | --- |
| Expert Title |  Kappa Statistics |
| IM | AG |
| Expert A | Expert B | 0.87 | 0.88 |
| Expert A | Expert C | 0.98 | 0.94 |
| Expert B | Expert C | 0.85 | 0.87 |

Table S3. The Accuracy of Different Frameworks of CNN model for Recognition of Atrophic gastritis and Gastric Intestinal Metaplasia

|  |  |
| --- | --- |
| Framework | Accuracy (%) |
| Atrophic Gastritis | Gastric Intestinal Metaplasia |
| TResNet | 96.36 | 97.60 |
| ResNet18 | 95.51 | 96.09 |
| ResNet34 | 93.25 | 95.03 |
| ResNet50 | 94.32 | 94.09 |
| DenseNet121 | 89.88 | 90.23 |
| VGG19 | 93.43 | 93.96 |
| VGG16 | 96.35 | 93.90 |

Table S4. Diagnosis performance between AI and experts

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | AI | Expert 1 | P-value | Expert 2 | P-value | Expert 3  | P-value |
| Atrophic Gastritis |
| Accuracy | 96.4% | 70.6% | <0.001 | 65.5% | <0.001 | 58.9% | <0.001 |
| Sensitivity | 96.2% | 51.7% | <0.001 | 35.2% | <0.001 | 36.8% | <0.001 |
| Specificity | 96.4% | 80.1% | <0.001 | 80.6% | <0.001 | 69.9% | <0.001 |
| Gastric Intestinal Metaplasia |
| Accuracy | 97.6% | 77.6% | <0.001 | 68.8% | <0.001 | 73.0% | <0.001 |
| Sensitivity | 97.9% | 42.0% | <0.001 | 28.2% | <0.001 | 47.3% | <0.001 |
| Specificity | 97.5% | 96.1% | 0.166 | 90.0% | <0.001 | 86.4% | <0.001 |

Table S5. Endoscopists vs AI in validation arm (N=274)

|  |  |  |
| --- | --- | --- |
| **Test dataset** | **Endoscopists** | **AI** |
| Accuracy | 67.2% (95%CI 61.3-72.7%) | 98.5% |
| Sensitivity | 41.9% (95%CI 33.8-50.3%) | 98.6% |
| Specificity | 96.8% (95%CI 92.1-99.1%) | 98.4% |