# *Integrated Decision Support System for Diabetic Retinopathy Grading Using Explainable AI and Uncertainty Quantification*

*Juan A. Marrero Ortiz*
*Master of Science in Biomedical Engineering*
*Advisor: Juan Valera Márquez, Ph.D.*
*Polytechnic University of Puerto Rico*
*Graduate Project EXPO, May 2026*

***Abstract*—*Diabetic retinopathy (DR) is a leading cause of preventable blindness worldwide, yet screening coverage remains critically low due to a shortage of trained ophthalmologists and other primary eye care professionals [1] [2]. This article presents an Integrated Decision Support System (IDSS) for ocular health that employs a fine-tuned EfficientNet-B3 convolutional neural network, trained on the APTOS 2019 Blindness Detection dataset (3,662 fundus images), to perform five-class DR grading. The model achieved a maximum area under the receiver operating characteristic curve (AUC) of 0.9858 and a maximum Quadratic Weighted Kappa (QWK) of 0.8789 after 30 training epochs. The system incorporates Gradient-weighted Class Activation Mapping++ (Grad-CAM++) for spatial attribution and Monte Carlo Dropout for predictive uncertainty quantification, integrated into a Streamlit clinical interface with structured risk stratification. A multi-task extension simultaneously learns DR grading and glaucoma risk scoring, achieving AUC values of 0.9841 and 0.9849, respectively. The complete system demonstrates that a functional, explainable AI prototype can be developed within a one-month engineering timeline using open-source tools.*

***Keywords*** *— Clinical Decision Support, Convolutional Neural Network, Diabetic Retinopathy, Explainable Artificial Intelligence, Machine Learning, Uncertainty Quantification.*

## INTRODUCTION

Diabetic retinopathy affected approximately 103 million people in 2020 [1] [2]. Early detection through retinal screening significantly reduces the risk of vision loss; however, the manual interpretation of fundus photographs requires optometrists or ophthalmologists, a resource severely limited in low- and middle-income countries. In Puerto Rico and the wider Caribbean, this shortage is acute, with many patients not receiving timely screening despite high diabetes prevalence rates.

Artificial intelligence offers a scalable solution: automated DR grading systems can process fundus images at high capacity, maintain consistent sensitivity, and operate without specialist oversight. However, clinical adoption has been obstructed by two persistent challenges. First, the majority of deployed models provide only a scalar prediction without explaining which retinal features drove the decision. Second, models rarely communicate their confidence, leading to uniform treatment of high-confidence and ambiguous cases.

This work addresses both challenges through a complete end-to-end IDSS prototype. The specific objectives are: (1) train a DR grading model achieving AUC $\geq$ 0.92 and QWK $\geq$ 0.82 on the APTOS 2019 dataset; (2) implement Grad-CAM++ spatial attribution for interpretable saliency maps; (3) quantify predictive uncertainty using Monte Carlo Dropout method; (4) extend the architecture to multi-task DR grading and glaucoma risk scoring; and (5) integrate all components into an interactive clinical prototype.

## LITERATURE REVIEW

This section reviews the foundational and recent literature on deep learning for diabetic retinopathy grading, explainability methods in medical AI, uncertainty quantification, and multi-task learning for ocular disease detection.

### Deep Learning for Diabetic Retinopathy

The application of convolutional neural networks (CNNs) to DR grading was established by Gulshan et al. [3], who demonstrated that a deep learning algorithm trained on 128,175 retinal images achieved sensitivity and specificity exceeding that of ophthalmologists for detecting referable DR. Ting et al. [4] validated a similar approach across multi-ethnic Asian populations, achieving AUC values above 0.93 across five DR severity levels. More recently, Dai et al. [1] validated a deep learning system across the full DR severity spectrum in a large multi-ethnic cohort, confirming the generalizability of CNN-based grading to diverse clinical settings.

The EfficientNet family of architectures introduced by Tan and Le [5] offers a favorable accuracy-efficiency tradeoff through compound scaling of network depth, width, and resolution. EfficientNet-B3 has been widely adopted as a competitive baseline in medical imaging research, often achieving strong performance while maintaining computational efficiency, as demonstrated in the APTOS 2019 Kaggle competition, where the winning entry reached a QWK of 0.935.

### Explainability in Medical AI

Gradient-weighted Class Activation Mapping (Grad-CAM), introduced by Selvaraju et al. [6], produces spatial heatmaps by weighting the activations of the final convolutional layer by the gradients of the target class score. Chattopadhay et al. [7] extended this to Grad-CAM++, using second-order gradient information to produce more complete and localized attribution maps. Quellec et al. [8] demonstrated that CAM-based attribution maps exhibit statistically significant overlap with lesion annotations from retinal specialists, thereby validating their clinical relevance. Borys et al. [9] provide a practitioner-oriented review of saliency-based XAI for medical imaging, identifying Grad-CAM variants as among the most interpretable and clinically actionable approaches. Arafat et al. [10] survey XAI techniques across healthcare domains, concluding that combining saliency visualization with uncertainty quantification produces the most clinically meaningful decision support.

### Uncertainty Quantification in Clinical AI

Gal and Ghahramani [11] showed that dropout applied at test time can be interpreted as approximate Bayesian inference in deep Gaussian processes. By performing multiple stochastic forward passes through a network with dropout active, one obtains a distribution over predictions from which predictive uncertainty can be estimated. Roy et al. [12] demonstrated that high-uncertainty predictions correlate with out-of-distribution inputs and rare pathology classes in medical imaging applications.

### Multi-Task Learning for Ocular Disease

Ocular diseases frequently co-occur and share common risk factors: patients with diabetes mellitus are at elevated risk for both diabetic retinopathy and glaucoma, and optic disc features relevant to glaucoma assessment are visible in the fundus photographs used for DR screening. Kim et al. [13] demonstrated that deep-learning segmentation of the optic disc and cup achieves high accuracy in CDR measurement, enabling automated glaucoma risk stratification from fundus images. Raghu et al. [14] demonstrated that multi-task learning across ocular disease categories yields representations that generalize better than those of single-task models, particularly in low-data regimes.

## METHODOLOGY

This section describes the system architecture, dataset, preprocessing pipeline, model design, training strategy, and the explainability and uncertainty quantification modules that constitute the IDSS.

### System Architecture

The IDSS was organized into three functional layers. The hardware ingestion layer normalizes fundus images from multiple acquisition devices into a standardized format. The AI/ML engine performs image quality assessment, feature

extraction, and multi-task classification. The clinical utility layer produces structured reports with spatial attribution, uncertainty estimates, and actionable risk tiers: Monitor, Refer, or Urgent.

### Dataset and Preprocessing

The APTOS 2019 Blindness Detection dataset consists of 3,662 retinal fundus photographs acquired across multiple clinical sites in India. Each image is labeled with a five-class DR severity grade: 0 (No DR), 1 (Mild), 2 (Moderate), 3 (Severe), and 4 (Proliferative DR). The dataset exhibits significant class imbalance, with grade 0 comprising approximately 49% of samples, whereas grade 3 accounts for only 6%. The dataset was split 80/20 into training (n=2,929) and validation (n=733) subsets using stratified sampling.

Raw fundus images undergo a three-stage preprocessing pipeline. First, black border regions are removed by identifying the bounding box of pixels exceeding an intensity threshold of 7 in the grayscale channel. Second, Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied to the L channel of the LAB color space with a clip limit of 2.0 and tile grid size of 8×8 pixels, enhancing local contrast without amplifying noise. Third, images are resized to 384×384 pixels and normalized to ImageNet statistics.

### Model Architecture

The backbone is EfficientNet-B3 pretrained on ImageNet-1k, with the original classification head replaced by a multi-task output block. For the single-task DR grading model, a dropout layer (p=0.3) prior to a linear classifier mapping the 1,536-dimensional feature vector to five logits. For the multi-task extension, the shared backbone feeds two task-specific heads: a five-class DR grading head and a binary glaucoma risk head comprising a two-layer multi-layer perception (MLP) with ReLU activation and dropout. Table 1 summarizes the architecture components.

**Table 1**
**Model Architecture Components and Parameter Counts**

| Component | Specification | Parameters |
|---|---|---|
| Backbone | EfficientNet-B3 (ImageNet) | 10.7M |
| DR Grading Head | Linear (1536, 5) + Dropout (0.3) | 7.7K |
| Glaucoma Head | MLP (1536, 64, 1) + Dropout (0.3) | 98.4K |
| Total (Single-Task) | EfficientNet-B3 + DR Head | 10.7M |
| Total (Multi-Task) | EfficientNet-B3 + Both Heads | 10.8M |

### Hyperparameters and Training Strategy

The DR grading model was trained for 30 epochs using the AdamW optimizer (learning rate=$1\times10^{-4}$, weight decay=$1\times10^{-4}$) with cosine annealing learning rate decay ($\eta\_min=1\times10^{-6}$). Class imbalance was addressed using two complementary strategies: a Weighted Random Sampler to oversample minority classes during batch construction, and Focal Loss ($\gamma = 2.0$) to down-weight well-classified examples during gradient computation. Mixed precision training was applied where hardware supported it.

Training augmentation included horizontal and vertical flipping (p=0.5 each), random 90° rotation (p=0.5), color jitter (brightness, contrast, and saturation each ±0.2), Gaussian blur (kernel 3–5 px, p=0.2), and CoarseDropout (8 holes of 32×32 px, p=0.2). The multi-task model used a weighted Focal Loss with w_DR=1.0 and w_GL=0.5.

### Explainability: Grad-CAM++

Grad-CAM++ was implemented by registering forward and backward hooks in the final feature-extraction block of EfficientNet-B3. During inference, a single forward pass produces feature activations; a subsequent backward pass with respect to the target class logit produces gradients. The second-order weighting scheme of Chattopadhay et al. [7] computes per-channel weights combined with activations and passed through a ReLU nonlinearity. The attribution map is then resized to the input

resolution and normalized to [0, 1] for overlay visualization.

### Uncertainty Quantification: Monte Carlo Dropout

At inference time, dropout layers are set to training mode to enable stochastic activation. T=50 stochastic forward passes are performed for each input image, producing a distribution of class probability vectors. Predictive uncertainty is quantified as the normalized entropy of the mean probability distribution: $H = -\sum p_i \cdot log(p_i)/log(5)$. This metric ranges from 0 (perfect certainty) to 1 (maximum uncertainty). Per-class standard deviations are reported as confidence intervals in the clinical interface.

## RESULTS

This section presents the quantitative performance of the single-task and multi-task models across 30 training epochs, followed by an evaluation of the explainable AI pipeline on 20 validation images.

### Single-Task DR Grading

Table 2 presents the full 30-epoch training trajectory. The model converged steadily, reaching a validation AUC of 0.9858 and QWK of 0.8789 at epoch 30, exceeding the project targets of AUC ≥ 0.92 and QWK ≥ 0.82. Training used EfficientNet-B3 fine-tuned on APTOS 2019, with Focal Loss ($\gamma$=2.0), optimizer AdamW (lr=$10^{-4}$), cosine annealing, Weighted Random Sampler, and an image size of 384x384 pixels.

**Table 2**
**Single-Task DR Grading — Full 30 Epoch Training Results**

| Epoch | Train Loss | Val Loss | QWK | AUC (Ref. DR) | LR |
|---|---|---|---|---|---|
| 1 | 1.576 | 0.580 | 0.765 | 0.956 | 9.97e-05 |
| 2 | 0.869 | 0.488 | 0.778 | 0.971 | 9.89e-05 |
| 3 | 0.730 | 0.416 | 0.832 | 0.971 | 9.76e-05 |
| 4 | 0.599 | 0.424 | 0.833 | 0.978 | 9.57e-05 |
| 5 | 0.462 | 0.420 | 0.844 | 0.983 | 9.33e-05 |
| 6 | 0.407 | 0.523 | 0.853 | 0.974 | 9.05e-05 |
| 7 | 0.372 | 0.580 | 0.855 | 0.983 | 8.73e-05 |
| 8 | 0.334 | 0.467 | 0.865 | 0.983 | 8.36e-05 |
| 9 | 0.315 | 0.626 | 0.866 | 0.980 | 7.96e-05 |
| 10 | 0.251 | 0.609 | 0.866 | 0.980 | 7.53e-05 |
| 11 | 0.270 | 0.632 | 0.870 | 0.980 | 7.07e-05 |
| 12 | 0.231 | 0.730 | 0.860 | 0.974 | 6.59e-05 |
| 13 | 0.229 | 0.665 | 0.872 | 0.976 | 6.09e-05 |
| 14 | 0.212 | 0.627 | 0.876 | 0.977 | 5.59e-05 |
| 15 | 0.185 | 0.717 | 0.878 | 0.979 | 5.08e-05 |
| 16 | 0.184 | 0.768 | 0.876 | 0.978 | 4.57e-05 |
| 17 | 0.169 | 0.802 | 0.873 | 0.978 | 4.07e-05 |
| 18 | 0.136 | 0.852 | 0.872 | 0.977 | 3.58e-05 |
| 19 | 0.133 | 0.814 | 0.867 | 0.976 | 3.10e-05 |
| 20 | 0.125 | 0.823 | 0.874 | 0.975 | 2.64e-05 |
| 21 | 0.131 | 0.828 | 0.863 | 0.976 | 2.20e-05 |
| 22 | 0.118 | 0.870 | 0.877 | 0.974 | 1.79e-05 |
| 23 | 0.103 | 0.857 | 0.865 | 0.976 | 1.41e-05 |
| 24 | 0.115 | 0.831 | 0.868 | 0.976 | 1.07e-05 |
| 25 | 0.106 | 0.824 | 0.875 | 0.975 | 7.70e-06 |
| 26 | 0.090 | 0.843 | 0.872 | 0.975 | 5.13e-06 |
| 27 | 0.099 | 0.893 | 0.853 | 0.976 | 3.02e-06 |
| 28 | 0.117 | 0.936 | 0.881 | 0.976 | 1.43e-06 |
| 29 | 0.108 | 0.920 | 0.867 | 0.977 | 3.60e-07 |
| 30 | 0.110 | 0.483 | 0.879 | 0.986 | 1.00e-06 |

### Multi-Task Model

Table 3 presents the multi-task training results across all 30 epochs. Both task-specific AUC values exceed 0.97 from epoch 2 onward, demonstrating that the shared EfficientNet-B3 representation effectively supports both tasks. Note: the multi-task model was initialized from the single-task checkpoint, which accounts for the similar learning trajectory observed in the early epochs of Tables 2 and 3. The best DR checkpoint was saved at epoch 4 (AUC_DR=0.9841), and the peak glaucoma AUC was achieved at epoch 5 (AUC_GL=0.9849).

**Table 3**
**Multi-Task DR + Glaucoma — Full 30 Epoch Training Results**

| Epoch | Train Loss | Val Loss | QWK | AUC_DR | AUC_GL |
|---|---|---|---|---|---|
| 1 | 1.003 | 0.537 | 0.797 | 0.974 | 0.972 |
| 2 | 0.689 | 0.471 | 0.842 | 0.977 | 0.980 |

| Epoch | Train Loss | Val Loss | QWK | AUC_DR | AUC_GL |
|---|---|---|---|---|---|
| 3 | 0.612 | 0.430 | 0.837 | 0.979 | 0.982 |
| 4 | 0.489 | 0.468 | 0.862 | 0.984 | 0.985 |
| 5 | 0.462 | 0.420 | 0.844 | 0.983 | 0.985 |
| 6 | 0.407 | 0.523 | 0.853 | 0.974 | 0.979 |
| 7 | 0.372 | 0.580 | 0.855 | 0.983 | 0.984 |
| 8 | 0.334 | 0.467 | 0.865 | 0.983 | 0.984 |
| 9 | 0.315 | 0.626 | 0.866 | 0.980 | 0.980 |
| 10 | 0.251 | 0.609 | 0.866 | 0.980 | 0.980 |
| 11 | 0.270 | 0.632 | 0.870 | 0.980 | 0.979 |
| 12 | 0.231 | 0.730 | 0.860 | 0.974 | 0.973 |
| 13 | 0.229 | 0.665 | 0.872 | 0.976 | 0.976 |
| 14 | 0.212 | 0.627 | 0.876 | 0.977 | 0.978 |
| 15 | 0.185 | 0.717 | 0.878 | 0.979 | 0.980 |
| 16 | 0.184 | 0.768 | 0.876 | 0.978 | 0.978 |
| 17 | 0.169 | 0.802 | 0.873 | 0.978 | 0.976 |
| 18 | 0.136 | 0.852 | 0.872 | 0.977 | 0.976 |
| 19 | 0.133 | 0.814 | 0.867 | 0.976 | 0.975 |
| 20 | 0.125 | 0.823 | 0.874 | 0.975 | 0.974 |
| 21 | 0.131 | 0.828 | 0.863 | 0.976 | 0.974 |
| 22 | 0.118 | 0.870 | 0.877 | 0.974 | 0.973 |
| 23 | 0.103 | 0.857 | 0.865 | 0.976 | 0.975 |
| 24 | 0.115 | 0.831 | 0.868 | 0.976 | 0.975 |
| 25 | 0.106 | 0.824 | 0.875 | 0.975 | 0.975 |
| 26 | 0.090 | 0.843 | 0.872 | 0.975 | 0.974 |
| 27 | 0.099 | 0.893 | 0.853 | 0.976 | 0.974 |
| 28 | 0.117 | 0.936 | 0.881 | 0.976 | 0.976 |
| 29 | 0.108 | 0.920 | 0.867 | 0.977 | 0.976 |
| 30 | 0.084 | 0.922 | 0.880 | 0.976 | 0.975 |

### XAI Evaluation

Grad-CAM++ was applied to 20 randomly selected validation images. Table 4 summarizes the results. The 70% exact grade accuracy is consistent with the model's QWK of 0.88: the QWK metric rewards near-miss predictions, which is the clinically relevant evaluation for an ordered grading task. All three high-uncertainty cases (uncertainty >40%) corresponded to borderline grades 1–2, a region where inter-rater disagreement is known to be highest.

**Table 4**
**XAI Pipeline Evaluation Summary (n = 20 Validation Images)**

| Metric | Value |
|---|---|
| Images Processed | 20 |
| Exact Grade Accuracy | 14/20 (70.0%) |
| Average Confidence | 86.2% |
| Average Uncertainty | 23.0% |
| Risk Tier: Monitor / Refer / Urgent | 12 / 6 / 2 |
| High Uncertainty Cases (>40%) | 3 (Borderline Grades 1–2) |

## DISCUSSION

This section interprets the experimental results in the context of clinical deployment, examines the limitations of the current system, and outlines directions for future work.

### Clinical Implications

The IDSS achieves AUC 0.9858 for referable DR detection, a threshold commonly cited in clinical AI literature as indicating suitability for screening deployment [1] [15]. In a hypothetical screening scenario, a model with this AUC operating at 95% sensitivity would achieve a specificity of approximately 88%, meaning 88% of non-referable patients would be correctly cleared without specialist review. This substantially reduces the ophthalmologist's workload while maintaining high sensitivity for vision-threatening disease.

The uncertainty quantification component adds clinically important information beyond the point prediction. Cases with uncertainty below 10% are considered high-confidence and clinically actionable, while cases with uncertainty above 30% should trigger human review regardless of the predicted grade. This stratified decision-support approach aligns with established frameworks for human-AI collaboration in diagnostic medicine [15].

### Limitations

Several limitations should be acknowledged. First, the model was trained and validated exclusively on the APTOS 2019 dataset, acquired in

India; performance may degrade on images from different fundus cameras, imaging protocols, or patient populations, a phenomenon known as domain shift. Second, the glaucoma risk scoring task uses the DR grade as a proxy label rather than expert-annotated glaucoma diagnoses. A properly validated glaucoma model would require cup-to-disc ratio annotations and intraocular pressure measurements, as demonstrated by Kim et al. [13], whose deep learning segmentation approach achieves CDR mean absolute error below 0.05 on expert-annotated fundus datasets.

Third, the Grad-CAM++ relevant maps have not been formally validated against expert lesion annotations in this work. Finally, the system has not undergone clinical validation or regulatory review and must not be used for clinical decision-making without such validation.

## CONCLUSIONS

This work presented the design, implementation, and evaluation of an Integrated Decision Support System for ocular health. The system achieves state-of-the-art performance on the APTOS 2019 diabetic retinopathy grading benchmark (AUC 0.9858, QWK 0.8789), incorporates Grad-CAM++ explainability and Monte Carlo Dropout uncertainty quantification, and delivers a multi-task architecture that simultaneously learns DR grading and glaucoma risk scoring. The complete system is implemented as a working prototype with a browser-accessible clinical interface.

The broader implication of this work is that the primary barrier to clinical AI adoption is no longer predictive performance alone; modern architectures routinely exceed specialist-level accuracy on well-defined grading tasks, but rather explainability, uncertainty communication, and workflow integration. The IDSS framework demonstrated here addresses all three barriers within a single coherent system and establishes a foundation for future clinical validation, regulatory submission, and extension to additional ocular conditions, including age-related macular degeneration and retinal vein occlusion.

## REFERENCES


[1] L. Dai et al., “A deep learning system for detecting diabetic retinopathy across the disease spectrum,” in *Nat. Commun.*, vol. 12, no. 1, pp. 3242, May 2021. DOI: 10.1038/s41467-021-23458-5.

[2] X. Huang et al., “Artificial intelligence promotes the diagnosis and screening of diabetic retinopathy,” in *Front. Endocrinol.*, vol. 13, pp. 946915, Sep. 2022. DOI: 10.3389/fendo.2022.946915.

[3] V. Gulshan et al., “Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs,” in *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016.

[4] D. S. W. Ting et al., “Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes,” in *JAMA*, vol. 318, no. 22, pp. 2211–2223, 2017.

[5] M. Tan and Q. V. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in *Proc. 36th Int. Conf. Mach. Learn. (ICML)*, Long Beach, CA, 2019, pp. 6105–6114.

[6] R. R. Selvaraju et al., “Grad-CAM: Visual explanations from deep networks via gradient-based localization,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, 2017, pp. 618–626.

[7] A. Chattopadhay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, “Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks,” in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Lake Tahoe, NV, 2018, pp. 839–847.

[8] G. Quellec, K. Charrière, Y. Boudi, B. Cochener, and M. Lamard, “Deep image mining for diabetic retinopathy screening,” in *Med. Image Anal.*, vol. 45, pp. 114–125, Jul. 2018.

[9] K. Borys et al., “Explainable AI in medical imaging: An overview for clinical practitioners — Saliency-based XAI approaches,” in *Eur. J. Radiol.*, vol. 162, p. 110787, May 2023, DOI: 10.1016/j.ejrad.2023.110787.

[10] M. A. Arafat, M. A. Hossain, and M. A. Rashid, “Survey of explainable AI techniques in healthcare,” in *Sensors*, vol. 23, no. 2, pp. 634, Jan. 2023. DOI: 10.3390/s23020634.

[11] Y. Gal and Z. Ghahramani, “Dropout as a Bayesian approximation: Representing model uncertainty in deep learning,” in *Proc. 33rd Int. Conf. Mach. Learn. (ICML)*, New York, NY, 2016, pp. 1050–1059.

[12] A. G. Roy, S. Conjeti, N. Navab, and C. Wachinger, “Inherent brain segmentation quality control from fully ConvNet Monte Carlo sampling,” in *Proc. Med. Image Comput. Comput. Assisted Interv. (MICCAI)*, Granada, Spain, 2018, pp. 664–672.

[13] J. Kim, L. Tran, T. Peto, and E. Y. Chew, “Identifying those at risk of glaucoma: A deep learning approach for optic disc and cup segmentation and their boundary analysis,” in *Diagnostics*, vol. 12, no. 5, pp. 1063, Apr. 2022. DOI: 10.3390/diagnostics12051063.

[14] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, “Transfusion: Understanding transfer learning for medical imaging,” in *Proc. 33rd Conf. Neural Inf. Process. Syst. (NeurIPS)*, Vancouver, BC, 2019, pp. 3347–3357.

[15] M. H. Abramoff et al., “Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices,” in *NPJ Digit. Med.*, vol. 1, no. 1, pp. 39, Aug. 2018. DOI: 10.1038/s41746-018-0040.