Appendix E. Potential Biases & Failure Modes

The system may exhibit performance limitations under real-world conditions that differ from the controlled training environment. In particular, domain shift can occur due to variations in lighting, camera angle, background clutter, or the presence of multiple foods within a single image. Low-light indoor or nighttime scenes may introduce noise, while extreme angles, close-ups, or partial crops can obscure key visual features. Background elements such as trays, hands, utensils, or menus may further distract the model. In such cases, the classifier may be biased toward the class with the most dominant visual pattern, such as noodles, broth, or fried textures. These patterns were qualitatively observed using evaluation-only images within a limited evidence scope and are addressed through capture guidance (e.g., bright lighting and top or 45-degree angles), manual override options, and planned future expansion by condition-specific data buckets.

Another inherent limitation arises from the classifier's label-space design. The image model is trained on 31 food classes, while the service-level Food DB covers approximately 300 foods. As a result, foods outside the trained label set may be predicted as the most visually similar among the 31 classes. This behavior reflects a label-space constraint rather than a functional bug. To mitigate this, the system allows users to bypass image recognition and continue the service flow through text-based database search and manual selection when recognition is uncertain.

Look-alike food categories also present a known source of confusion, particularly among visually similar items such as noodle or soup-based dishes (e.g., ramen, kalguksu, jjajangmyeon, jjamppong), fried foods (e.g., fried chicken versus yangnyeom chicken), and rice bowl variants (e.g., kimchi fried rice versus bibimbap). These confusion patterns were observed in evaluation-only settings and directly informed the design of user-correction mechanisms, including the potential presentation of top-k candidates and explicit user selection or correction in the UI.

Nutritional information introduces additional uncertainty due to natural variation across brands, recipes, and portion sizes. The Food DB therefore provides standardized "guide values" per serving, which may differ from a user's actual meal. To maintain transparency, serving units are explicitly disclosed, and representative provenance samples are documented. Users are encouraged to interpret nutritional values as approximate references, with portion adjustment options and future plans for expanded provenance and standardization.

Finally, label noise and ambiguity are especially relevant for evaluation-only internet-sourced images. Such images may lack clear context or depict mixed plates, resulting in weak or ambiguous ground truth. To address this, only qualitatively clear cases are considered for failure-mode observation, and evaluation-only data are not used to support quantitative accuracy claims. Instead, these observations are limited to identifying failure patterns and informing UX and mitigation strategies.