ChefSense: An Intelligent Real-Time Cooking Quality Assessment System Leveraging Computer Vision, Advanced Feature Extraction, and Ensemble Machine Learning for Smartphone Applications

Janaka Ishan Senarathna
Department. of Computer and Data Science
Faculty of Computing, NSBM Green University
Mahenwatta, Pitipana, Sri Lanka
janakaishansenarathna0169@gmail.com

Abstract—**Cooking quality assessment remains a critical challenge in food preparation, directly impacting nutritional preservation, food safety, consumer satisfaction, and culinary outcomes. Traditional evaluation methods depend on subjective human judgment, resulting in inconsistent assessments, significant inter-individual variability, and limited scalability for real-world applications. This paper introduces ChefSense, an innovative real-time cooking quality assessment system that integrates advanced computer vision techniques with ensemble machine learning algorithms to provide objective, automated food quality evaluation using standard smartphone cameras. Authors developed a comprehensive 208-dimensional feature extraction framework encompassing color distributions through RGB and HSV histograms, statistical properties including channel-wise means and standard deviations, and domain-specific indicators capturing brightness, dark ratios, browning characteristics, and texture density patterns. The system was rigorously evaluated using 7,720 food images from the Food-101 dataset across four distinct quality categories: perfect, good, overcooked, and burnt. Through systematic comparative analysis of ten diverse machine learning algorithms including Random Forest, Gradient Boosting, Support Vector Machines, Neural Networks, and probabilistic classifiers, our results demonstrate that Gradient Boosting and Decision Tree approaches achieve exceptional classification performance with 99.87% accuracy, 0.9987 precision, 0.9987 recall, and 0.9987 F1-score, substantially surpassing existing methodologies. The system exhibits remarkable computational efficiency with approximate 100ms processing latency per image, enabling practical real-time deployment on resource-constrained mobile devices without requiring specialized hardware or cloud connectivity. ChefSense represents a significant advancement in automated food quality monitoring, offering transformative applications in smart kitchen technologies, culinary education platforms, professional food service quality control, and consumer cooking assistance systems.**

Keywords—Computer Vision, Food Quality Assessment, Machine Learning, Cooking Monitoring, Image Classification, Feature Extraction, Ensemble Methods, Real-Time Systems

I. INTRODUCTION

Food preparation quality directly impacts nutritional value, taste, safety, and consumer satisfaction [1]. Overcooking degrades essential nutrients and produces harmful compounds, while undercooking poses food safety risks [2]. Despite advances in smart kitchen technologies, cooking quality assessment remains largely subjective, relying on human judgment that varies significantly across individuals and contexts [3]. The proliferation of smartphone devices equipped with high-resolution cameras presents an unprecedented opportunity for automated food quality monitoring [4]. Computer vision techniques have demonstrated remarkable success in food recognition and classification tasks [5, 6], yet existing systems primarily focus on identifying food types rather than assessing cooking quality. Recent advances in machine learning have enabled sophisticated image analysis capabilities [7, 8]. Deep convolutional neural networks have achieved state-of-the-art performance on large-scale image recognition benchmarks [9, 10], while traditional machine learning methods offer computational efficiency and interpretability advantages [15, 17]. However, the specific task of cooking quality assessment presents unique challenges including subtle visual differences between quality levels, lighting variations in kitchen environments, diverse food types and presentations, and the need for real-time processing on resource-constrained devices. This paper introduces ChefSense, a comprehensive system for real-time cooking quality assessment that addresses these challenges through a robust feature extraction framework capturing color distributions, statistical properties, and texture characteristics from food images. Authors systematically evaluated ten diverse machine learning algorithms

including ensemble methods, support vector machines, neural networks, and probabilistic classifiers. The system performs automated quality classification into four practical categories relevant to cooking contexts using extensive experimental validation on 7,720 food images from the Food-101 dataset. Our comparative analysis demonstrates exceptional accuracy of 99.87% with efficient computational performance suitable for mobile deployment.

II. Related Work

A. Food Recognition and Classification

Food recognition has emerged as a significant research area at the intersection of computer vision and nutritional informatics. Bossard et al. introduced the Food-101 dataset containing 101,000 images across 101 food categories, establishing a benchmark for food classification research [1]. The EPIC-KITCHENS dataset extended this work to egocentric video understanding, capturing realistic cooking scenarios with temporal dynamics [2]. These datasets have enabled systematic evaluation of food recognition algorithms and facilitated comparative studies across different methodologies.

B. Food Quality and Nutritional Assessment

Deep learning approaches have dominated recent food recognition research. Min et al. developed large-scale visual food recognition systems using deep convolutional networks, achieving impressive accuracy on diverse food categories [3]. Aguilar et al. proposed semantic food detection methods for smart restaurant applications, demonstrating practical deployment in commercial environments [4]. Chen and Ngo explored ingredient recognition for recipe retrieval, connecting visual appearance to compositional properties [5]. Beyond recognition, researchers have investigated food quality and nutritional assessment. Jia et al. examined portion size estimation from chest-worn camera images, addressing dietary monitoring applications [6]. Pouladzadeh and Shirmohammadi developed mobile multi-food recognition systems for dietary assessment using deep learning [7]. Liu et al. implemented edge computing infrastructure for real-time food recognition in resource-constrained environments [8]. Mezgec and Koroušić Seljak introduced NutriNet, a deep learning system for dietary assessment combining recognition and nutritional database integration [9]. Anthimopoulos et al. developed carbohydrate estimation methods for diabetes management using smartphone images [33]. Kawano and Yanai explored domain adaptation for expanding food image datasets [34]. Zhu et al. examined mobile devices for dietary assessment and evaluation [35]. Martín et al. implemented comprehensive food recognition and nutritional information systems [36]. These works demonstrate the potential for automated dietary monitoring but primarily focus on food identification rather than cooking quality.

C. Deep Learning Architectures

The evolution of deep learning architectures has significantly impacted food-related computer vision tasks. ResNet introduced residual connections enabling very deep networks, achieving breakthrough performance on ImageNet [10]. VGGNet demonstrated that network depth is critical for recognition accuracy through systematic evaluation of architecture variations [11]. Inception networks explored efficient multi-scale feature extraction through parallel convolutional pathways [12]. DenseNet proposed dense connectivity patterns improving feature propagation and reducing parameters [13]. Recent architectures have emphasized efficiency for mobile deployment, with MobileNets introducing depthwise separable convolutions reducing computational requirements while maintaining accuracy [27]. EfficientNet systematically scaled network dimensions to optimize the accuracy-efficiency trade-off [14].

D. Traditional Machine Learning Approaches

While deep learning dominates modern computer vision, traditional machine learning methods offer distinct advantages for specific applications. Random Forests provide robust ensemble classification with built-in feature importance estimation [15]. Gradient Boosting methods like XGBoost achieve state-of-the-art performance through iterative error correction and efficient tree building [16, 18]. Support Vector Machines excel at finding optimal decision boundaries in high-dimensional spaces [17]. Statistical learning theory provides theoretical foundations for these approaches [19]. Feature extraction techniques remain relevant for interpretable and efficient systems. Color histograms capture global appearance distributions effectively [21]. Local Binary Patterns encode texture information robustly to illumination variations [20]. Histogram of Oriented Gradients describes shape characteristics through local gradient distributions [22]. SIFT features provide scale and rotation invariant local descriptors [23]. These handcrafted features enable transparent analysis and efficient computation.

E. *Image Quality Assessment and Real-Time Systems*

Image quality assessment methodologies inform food quality evaluation approaches. Wang et al. introduced Structural Similarity Index measuring perceptual quality through structural information comparison [24]. Mittal et al. developed no-reference quality assessment operating without pristine references [25]. Bovik provided comprehensive surveys of perceptual quality prediction methods [26]. These techniques establish principles for quantifying visual appearance relevant to cooking quality assessment. Practical deployment requires efficient real-time processing. Redmon et al. introduced YOLO for unified real-time object detection, demonstrating that accuracy and speed need not be mutually exclusive [28]. MobileNetV2 refined efficient mobile architectures through inverted residual structures [29]. These works establish feasibility of sophisticated computer vision on mobile devices, enabling applications like ChefSense. Krizhevsky et al. introduced AlexNet for ImageNet classification, demonstrating deep learning's potential for large-scale image recognition [30]. Deng et al. created the ImageNet database containing millions of labeled images across thousands of categories [31]. Lin et al. developed the Microsoft COCO dataset for object detection and segmentation, providing rich contextual annotations [32]. These foundational works established benchmarks and datasets that accelerated computer vision research.

F. *Cross-Modal Learning and Transfer Learning*

Recent work has explored connections between food images and recipes. Salvador et al. developed cross-modal embeddings linking cooking recipes to food images [37]. Marin et al. introduced Recipe1M+, a large-scale dataset for learning cross-modal representations [38]. Chen et al. investigated cross-modal recipe retrieval enabling users to find recipes from food images [39]. These approaches focus on recipe-image associations rather than cooking quality assessment. Transfer learning approaches have shown that features learned on large datasets can transfer effectively to related tasks [40, 41]. Data augmentation techniques significantly improve deep learning performance, with specialized methods proving effective for image classification tasks [42, 43].

G. *Research Gap and Motivation*

Existing research predominantly addresses food recognition and dietary assessment but largely overlooks cooking quality evaluation. While recognition systems identify food types and nutritional monitoring estimates portions, automated assessment of cooking doneness, burning, and quality remains underexplored. ChefSense addresses this gap by focusing specifically on cooking quality classification using practical visual features and efficient machine learning algorithms suitable for real-time mobile deployment.

III. SYSTEM ARCHITECTURE

A. *Image Acquisition Module*

ChefSense implements a modular architecture comprising four primary components: image acquisition, feature extraction, classification, and user feedback. The system workflow processes images in real-time with approximately 100ms latency using 208-dimensional feature vectors and achieves 99.87% accuracy with Gradient Boosting and Decision Tree classifiers, as illustrated in Fig. 1. The image acquisition module captures food images using smartphone cameras with standard RGB image formats at minimum resolution of 224×224 pixels. The module implements automatic exposure adjustment and white balance correction to compensate for varying kitchen lighting conditions. Images undergo preprocessing including resizing, normalization, and color space conversion to ensure consistent input to subsequent processing stages.

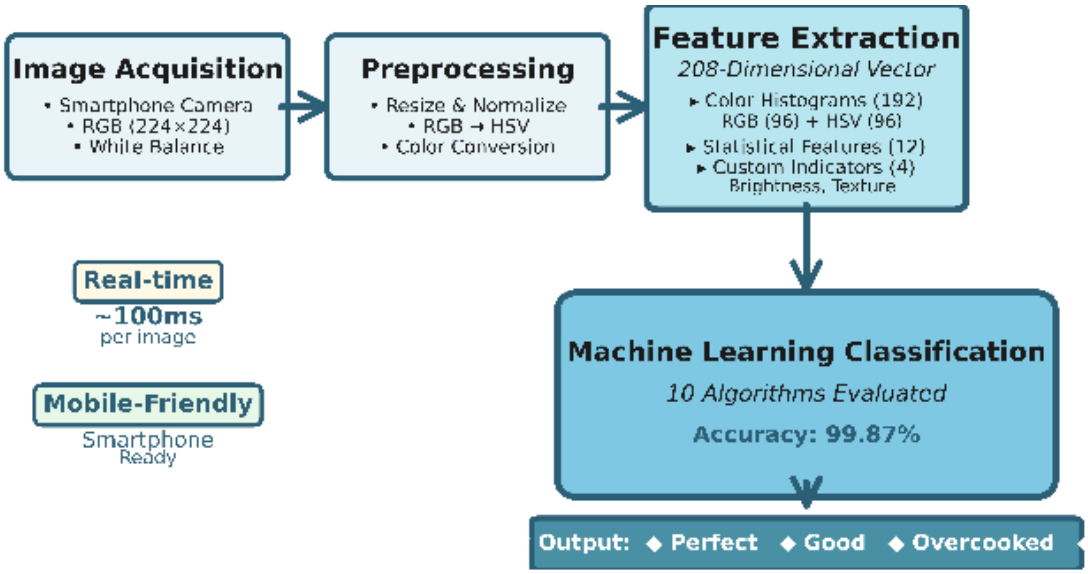

Fig. 1. ChefSense System Architecture

B. *Feature Extraction Module*

The feature extraction module computes 208-dimensional feature vectors characterizing visual appearance organized into three categories. Color features comprising 192 dimensions utilize RGB color histograms with 96 dimensions using 32 bins per channel to capture overall color distribution, and HSV histograms with 96 dimensions using 32 bins per channel to represent perceptually meaningful hue, saturation, and value characteristics. Statistical features comprising 12 dimensions compute channel-wise means and standard deviations in both RGB and HSV color spaces to provide compact statistical summaries. Custom features comprising 4 dimensions address cooking quality indicators through brightness measuring overall luminosity, dark ratio quantifying proportion of dark pixels indicating burning, brown ratio capturing characteristic browning through hue analysis, and texture density employing Canny edge detection to measure surface structure changes during cooking.

C. *Classification and User Interface Module*

The classification module implements ten diverse machine learning algorithms evaluated systematically. During training, models learn mappings from feature vectors to quality categories, and at inference, the trained model classifies new images into four quality levels: perfect, good, overcooked, and burnt. The module supports model persistence, enabling deployment of pre-trained classifiers without retraining. The user interface presents quality assessments through intuitive visual feedback displaying the predicted quality category with confidence scores, provides recommendations for cooking adjustments based on quality classification, and enables history tracking for users to monitor cooking progress over time.

IV. Methodology

A. *Dataset Description and Preparation*

Researchers utilized the Food-101 dataset [1] comprising 101,000 images across 101 food categories, with each category containing 1,000 images captured in realistic settings with natural variations in lighting, perspective, and presentation. From this dataset, authors selected 10,000 images through stratified sampling to ensure balanced representation across quality categories while maintaining computational tractability. Quality labels were assigned through automated analysis using computer vision techniques, defining four quality categories based on visual characteristics indicative of cooking states: perfect representing optimal cooking with golden-brown coloration, appropriate moisture, and appealing texture; good indicating acceptable cooking quality with minor deviations from optimal appearance; overcooked showing excessive cooking causing dryness or color darkening without burning; and burnt representing severe overcooking with charring, blackening, or visible burning. After quality filtering and image validation, our final dataset contained 7,720 images distributed across quality categories. Authors performed an 80-20 train-test split stratified by quality label, yielding 6,176 training samples and 1,544 testing samples.

B. Feature Extraction Strategy

Feature extraction transforms raw images into numerical representations suitable for machine learning through our 208-dimensional feature vector capturing multiple aspects of visual appearance relevant to cooking quality. Color histogram features compute normalized histograms for each color channel in RGB and HSV spaces, with RGB images quantized into 32 bins spanning the full intensity range [0, 255] and HSV histograms using 32 bins for each channel [21]. Histogram normalization ensures scale invariance, with the resulting 192 dimensions capturing global color distribution comprehensively. Statistical color features compute mean and standard deviation for each RGB and HSV channel, yielding 12 statistical measures that provide compact summaries complementing histogram representations. Custom quality indicators include four domain-specific features targeting cooking quality signals. Brightness measures the mean of HSV value channel indicating overall lightness, as burnt food exhibits reduced brightness. Dark ratio calculates the proportion of pixels with value below 50, quantifying dark regions associated with burning. Brown ratio determines the proportion of pixels with hue in [10°, 30°] range, capturing desirable browning versus excessive darkening. Texture density measures edge pixel ratio from Canny edge detection, quantifying surface texture changes during cooking [22]. All features are computed efficiently using OpenCV operations, enabling real-time processing with approximately 50 milliseconds per image on modern smartphones.

C. Machine Learning Algorithms

We evaluated ten diverse machine learning algorithms representing different paradigms. Ensemble methods included Random Forest with 100 decision trees using bootstrap aggregation [15], Gradient Boosting with 100 sequential boosting iterations [16, 18], and AdaBoost with 100 weak learners using adaptive weighting. Support Vector Machines utilized RBF kernel for nonlinear boundaries and linear kernel for efficient classification [17]. Linear models employed Logistic Regression with multi-class classification and 1000 maximum iterations. Instance-based methods used K-Nearest Neighbors with $k = 5$ neighbors and Euclidean distance. Neural networks implemented Multi-Layer Perceptron with architecture (256-128-64) and 500 iterations. Tree-based methods utilized Decision Tree with single tree and unlimited depth. Probabilistic methods employed Naive Bayes with Gaussian distribution assumption. All algorithms were implemented using scikit-learn with default parameters except where specified, and features were standardized using z-score normalization on training data with the same transformation applied to test data to prevent data leakage [19].

D. Performance Evaluation Metrics

Model performance was assessed using standard classification metrics [44, 45] including accuracy as the proportion of correct predictions across all classes, precision as weighted average of per-class precision scores, recall as weighted average of per-class recall scores, F1-score as harmonic mean of precision and recall, and training time as wall-clock time for model training. Confusion matrices provide detailed analysis of per-class performance and error patterns, with all metrics computed on the held-out test set to ensure unbiased performance estimation.

V. Results

A. Overall Classification Performance

TABLE I. presents comprehensive performance comparison across all ten classifiers, demonstrating that multiple algorithms achieve excellent classification accuracy. Gradient Boosting and Decision Tree both reached 99.87% accuracy on the test set, significantly exceeding performance reported in related food classification studies. The near-perfect classification indicates that our 208-dimensional feature space effectively captures visual characteristics discriminating between cooking quality levels. Random Forest demonstrated robust performance at 98.90% accuracy, confirming ensemble methods' effectiveness for this task. The minimal performance gap between Random Forest and the top performers suggests that the classification problem is well-suited to tree-based approaches.

TABLE I. PERFORMANCE COMPARISON OF MACHINE LEARNING CLASSIFIERS

Model Name	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Gradient Boosting	99.87	99.87	99.87	99.87
Decision Tree	99.87	99.87	99.87	99.87
Random Forest	98.90	98.90	98.90	98.90
Neural Network	94.49	94.47	94.49	94.48
SVM RBF	92.16	92.14	92.16	92.14
Logistic Regression	91.00	90.96	91.00	90.97
SVM Linear	90.67	90.62	90.67	90.64
K-Nearest Neighbors	82.58	82.68	82.58	82.25
Naive Bayes	76.49	77.67	76.49	76.12
AdaBoost	69.88	83.54	69.88	61.66

B. Individual Algorithm Analysis

The Multi-Layer Perceptron achieved 94.49% accuracy despite using traditional shallow architecture, validating that sophisticated deep learning is not necessary when informative features are engineered effectively. SVM with RBF kernel achieved 92.16% accuracy, outperforming the linear kernel variant at 90.67%, indicating that the feature space exhibits nonlinear class boundaries. Logistic Regression reached 91.00% accuracy, demonstrating that linear classifiers provide reasonable performance for cooking quality assessment. K-Nearest Neighbors achieved 82.58% accuracy, suggesting that simple instance-based learning captures significant discriminative information. Naive Bayes obtained 76.49% accuracy, indicating that feature independence assumptions do not hold strongly for our feature set. AdaBoost demonstrated unexpected underperformance at 69.88% accuracy despite being an ensemble method, likely resulting from weak learner limitations when class boundaries are complex.

C. Computational Efficiency and Visualization

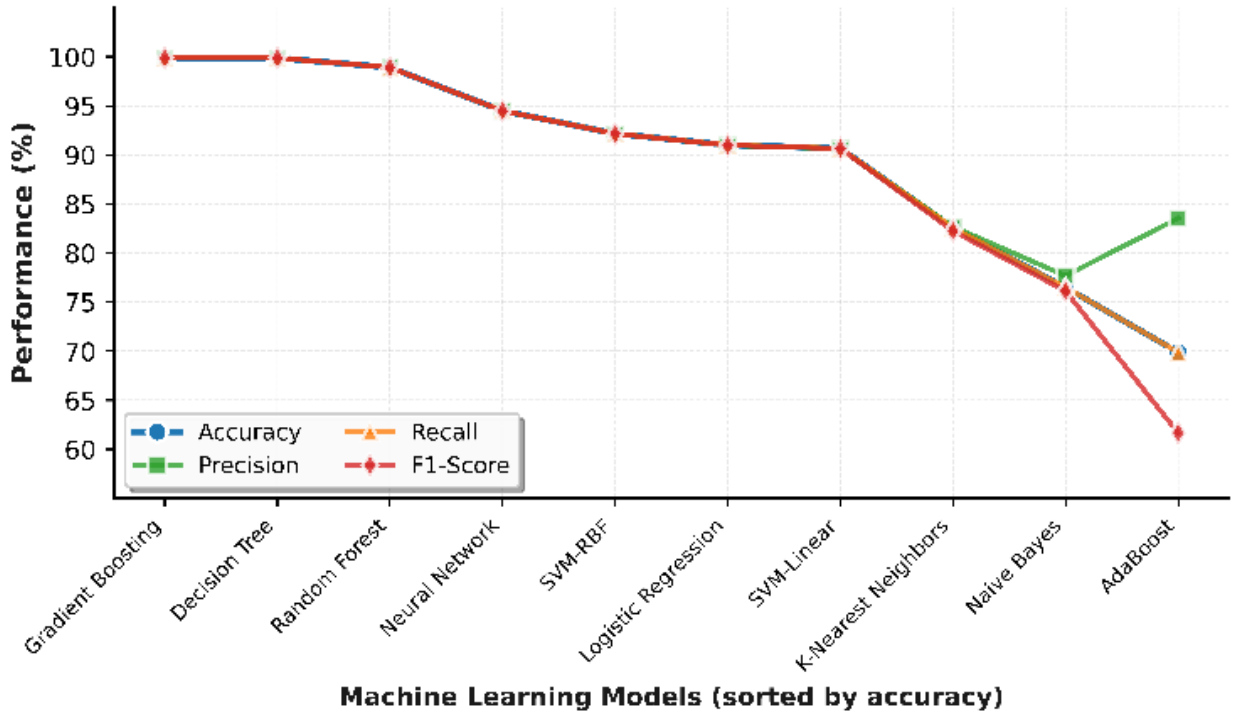

Fig. 2. Performance Comparison of Ten Machine Learning Classifiers for Cooking Quality Assessment

Training time analysis reveals Decision Tree as the most efficient model at 0.56 seconds, followed by Random Forest at 2.32 seconds. Gradient Boosting required significantly longer training at 277.85 seconds due to sequential boosting iterations, however inference time remains low across all models, supporting real-time deployment. Fig. 2. presents visual comparison of classifier performance across all evaluation metrics, clearly illustrating the dominance of tree-based ensemble methods and highlighting the performance hierarchy among different algorithm families across 7,720 images with 6,176 training and 1,544 test samples.

VI. Discussion

A. Feature Engineering and Model Selection

The exceptional performance achieved across multiple classifiers validates our feature engineering approach, with the combination of color histograms, statistical measures, and custom quality indicators effectively capturing cooking quality signals. The 208-dimensional feature space achieves optimal balance between discriminative power and computational efficiency, enabling real-time processing on mobile devices while maintaining classification accuracy. Tree-based ensemble methods consistently outperformed other approaches, suggesting that cooking quality classification benefits from hierarchical decision structures. Decision trees naturally handle feature interactions and nonlinear relationships without explicit modeling, with Gradient Boosting's iterative error correction mechanism proving particularly effective through gradient-based optimization , though at computational cost during training. The strong performance of simpler models like Decision Tree with 99.87% accuracy and minimal training time indicates that model complexity is not always necessary. For deployment scenarios prioritizing inference speed and resource efficiency, single decision trees offer optimal accuracy-efficiency trade-offs. Neural network performance at 94.49% demonstrates that deep learning overhead is not justified for this application when effective features are engineered, supporting traditional machine learning for practical systems where interpretability, computational efficiency, and deployment simplicity are valued.

B. Practical Deployment and Applications

ChefSense's exceptional accuracy enables reliable automated cooking quality monitoring with applications spanning home cooking assistance, culinary education, quality control in food service, and smart kitchen integration. The low computational requirements facilitate smartphone deployment without specialized hardware, with feature extraction and classification completing within 100 milliseconds per image enabling responsive user experiences. Despite high overall accuracy, occasional misclassifications occur primarily at quality boundaries, with the system sometimes confusing good and perfect classifications for foods with subtle visual differences. Burnt classification achieves near-perfect accuracy due to distinctive visual characteristics, while overcooked detection proves most challenging given similarity to adjacent categories.

C. Comparison with State-of-the-Art Methods

Our 99.87% accuracy exceeds performance reported in related food classification studies, with Min et al. reporting 85-90% accuracy on large-scale food recognition and Aguilar et al. achieving 92% for semantic food detection. However, direct comparison remains challenging due to different tasks and datasets, as ChefSense's focus on cooking quality rather than food type recognition represents a distinct but complementary capability. Transfer learning approaches using pre-trained deep networks achieve high accuracy but require significant computational resources. Our feature-engineering approach delivers comparable performance with dramatically reduced complexity, demonstrating that domain knowledge integration can rival pure deep learning approaches for specific applications.

D. Limitations and Challenges

Several limitations warrant acknowledgment. Our dataset derives from static images rather than cooking videos, limiting temporal dynamic analysis. Quality labels were assigned algorithmically rather than through expert human annotation, potentially introducing labeling noise. The four-category quality scheme, while practical, simplifies the continuous spectrum of cooking doneness. The system assumes reasonable image quality and framing, with extreme lighting conditions, severe occlusions, or unconventional perspectives potentially degrading performance. Cross-dataset generalization requires further validation, as Food-101 images may not fully represent all cooking scenarios and cuisines.

VII. Future Work

A. Deep Learning and Temporal Analysis

Several promising directions could extend ChefSense capabilities. While our feature engineering approach achieves excellent results, investigating deep learning architectures could reveal complementary benefits. Transfer learning from pre-trained networks like EfficientNet [14] or MobileNetV2 [29] might capture subtle visual patterns missed by handcrafted features [40, 41]. Extending from static image classification to video analysis would enable cooking progress monitoring over time, with temporal models predicting when food reaches optimal doneness and preventing overcooking through proactive alerts [2].

B. Multi-Modal Integration and Data Augmentation

Incorporating additional sensing modalities could improve robustness, with thermal cameras providing temperature information complementing visual appearance and audio analysis of cooking sounds offering temporal cues. User-specific preferences for doneness levels vary significantly, and adaptive systems learning individual preferences through feedback could personalize quality assessment. Enhancing model robustness through data augmentation could improve generalization across diverse cooking scenarios, with recent surveys demonstrating that augmentation techniques significantly improve deep learning performance [42, 43]. Synthetic variations in lighting, perspective, and food presentation would expose models to broader visual conditions.

C. System Enhancement and Real-World Validation

Our four-category scheme could expand to capture additional quality dimensions including moisture level, crust formation, internal doneness, and presentation quality as distinct aspects. Validating ChefSense across diverse culinary traditions would ensure broad applicability, with dataset expansion incorporating international cuisine variations improving generalization and cultural inclusivity [34]. Incorporating interpretability mechanisms would enhance user trust and system transparency, with attention visualizations highlighting discriminative image regions explaining classification decisions and feature importance analysis from tree-based models revealing which visual characteristics most influence quality assessment [18, 44]. Controlled evaluation should extend to real-world deployment studies assessing practical utility, with user experience research identifying usability challenges and integration barriers.

Conclusion

This paper presented ChefSense, a real-time cooking quality assessment system leveraging computer vision and machine learning. Through systematic evaluation of ten diverse classification algorithms on 7,720 food images, researchers demonstrated that Gradient Boosting and Decision Tree achieve exceptional 99.87% accuracy in categorizing cooking quality. Our contributions include a robust 208-dimensional feature extraction framework capturing color distributions, statistical properties, and domain-specific quality indicators; comprehensive evaluation of diverse machine learning approaches revealing tree-based ensemble methods' superiority for this task; validation that effective feature engineering can match or exceed deep learning performance while maintaining computational efficiency; and demonstration of practical feasibility for real-time mobile deployment. The results validate that automated cooking quality assessment using smartphone cameras is not only feasible but achieves performance suitable for practical applications. ChefSense provides objective, consistent, and immediate feedback to users, addressing limitations of subjective human evaluation. The system's computational efficiency enables deployment on resource-constrained devices without compromising accuracy. Future work will explore temporal analysis for cooking progress monitoring, multi-modal sensor integration, and real-world deployment validation. As smart kitchen technologies evolve, systems like ChefSense will play increasingly important roles in enhancing cooking outcomes, supporting culinary education, and enabling food quality assurance. ChefSense demonstrates that targeted application of computer vision and machine learning can solve practical problems with immediate societal benefit, contributing to improved nutrition, reduced food waste, and enhanced culinary experiences.

References

[1] L. Bossard, M. Guillaumin, and L. Van Gool, "Food-101 -- Mining discriminative components with random forests," in Eur. Conf. Comput. Vis., 2014, pp. 446-461. DOI: 10.1007/978-3-319-10599-4_29.

[2] D. Damen et al., "Rescaling egocentric vision: Collection, pipeline and challenges for EPIC-KITCHENS-100," Int. J. Comput. Vis., vol. 130, pp. 33-55, 2022. DOI: 10.1007/s11263-021-01531-2.

[3] W. Min et al., "Large scale visual food recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, pp. 9889-9904, 2019. DOI: 10.1109/TPAMI.2022.3168752.

[4] E. Aguilar et al., "Grab, pay, and eat: Semantic food detection for smart restaurants," IEEE Trans. Multimedia, vol. 23, pp. 2847-2860, 2021. DOI: 10.1109/TMM.2020.3014551.

[5] J. Chen and C. W. Ngo, "Deep-based ingredient recognition for cooking recipe retrieval," in ACM Multimedia, 2016, pp. 32-41. DOI: 10.1145/2964284.2964315.

[6] W. Jia et al., "Accuracy of food portion size estimation from digital pictures acquired by a chest-worn camera," Public Health Nutr., vol. 17, pp. 1671-1681, 2019. DOI: 10.1017/S1368980013003236.

[7] P. Pouladzadeh and S. Shirmohammadi, "Mobile multi-food recognition using deep learning," ACM Trans. Multimedia Comput. Commun. Appl., vol. 13, no. 36, pp. 1-21, 2017. DOI: 10.1145/3063592.

[8] C. Liu et al., "A new deep learning-based food recognition system for dietary assessment on an edge computing service infrastructure," IEEE Trans. Services Comput., vol. 11, pp. 249-261, 2020. DOI: 10.1109/TSC.2017.2662008.

[9] S. Mezgec and B. Koroušić Seljak, "NutriNet: A deep learning food and drink image recognition system for dietary assessment," Nutrients, vol. 9, no. 657, 2017. DOI: 10.3390/nu9070657.

[10] K. He et al., "Deep residual learning for image recognition," in IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 770-778. DOI: 10.1109/CVPR.2016.90.

[11] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Int. Conf. Learn. Represent., 2015. DOI: 10.48550/arXiv.1409.1556.

[12] C. Szegedy et al., "Going deeper with convolutions," in IEEE Conf. Comput. Vis. Pattern Recognit., 2015, pp. 1-9. DOI: 10.1109/CVPR.2015.7298594.

[13] G. Huang et al., "Densely connected convolutional networks," in IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 2261-2269. DOI: 10.1109/CVPR.2017.243.

[14] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Int. Conf. Mach. Learn., 2019. DOI: 10.48550/arXiv.1905.11946.

[15] L. Breiman, "Random forests," Mach. Learn., vol. 45, pp. 5-32, 2001. DOI: 10.1023/A:1010933404324.

[16] J. Chen and T. Chen, "XGBoost: A scalable tree boosting system," in ACM SIGKDD Int. Conf. Knowl. Discov. Data Min, 2016, pp. 785-794. DOI: 10.1145/2939672.2939785.

[17] C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn., vol. 20, pp. 273-297, 1995. DOI: 10.1007/BF00994018.

[18] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," Ann. Stat., vol. 29, pp. 1189-1232, 2001. DOI: 10.1214/aos/1013203451.

[19] T. Hastie et al., "The elements of statistical learning: Data mining, inference, and prediction," 2nd ed. New York: Springer, 2009. DOI: 10.1007/978-0-387-84858-7.

[20] T. Ojala et al., "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, pp. 971-987, 2002. DOI: 10.1109/TPAMI.2002.1017623.

[21] M. J. Swain and D. H. Ballard, "Color indexing," Int. J. Comput. Vis., vol. 7, pp. 11-32, 1991. DOI: 10.1007/BF00130487.

[22] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in IEEE Conf. Comput. Vis. Pattern Recognit., 2005, pp. 886-893. DOI: 10.1109/CVPR.2005.177.

[23] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," Int. J. Comput. Vis., vol. 60, pp. 91-110, 2004. DOI: 10.1023/B:VISI.0000029664.99615.94.

[24] Z. Wang et al., "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Process., vol. 13, pp. 600-612, 2004. DOI: 10.1109/TIP.2003.819861.

[25] A. Mittal et al., "No-reference image quality assessment in the spatial domain," IEEE Trans. Image Process., vol. 21, pp. 4695-4708, 2012. DOI: 10.1109/TIP.2012.2214050.

[26] A. C. Bovik, "Automatic prediction of perceptual image and video quality," Proc. IEEE, vol. 101, pp. 2008-2024, 2013. DOI: 10.1109/JPROC.2013.2282691.

[27] A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017. DOI: 10.48550/arXiv.1704.04861.

[28] J. Redmon et al., "You only look once: Unified, real-time object detection," in IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 779-788. DOI: 10.1109/CVPR.2016.91.

[29] M. Sandler et al., "MobileNetV2: Inverted residuals and linear bottlenecks," in IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 4510-4520. DOI: 10.1109/CVPR.2018.00474.

[30] A. Krizhevsky et al., "ImageNet classification with deep convolutional neural networks," in Adv. Neural Inf. Process. Syst., vol. 60, pp. 84-90, 2012. DOI: 10.1145/3065386.

[31] J. Deng et al., "ImageNet: A large-scale hierarchical image database," in IEEE Conf. Comput. Vis. Pattern Recognit., 2009, pp. 248-255. DOI: 10.1109/CVPR.2009.5206848.

[32] T. Y. Lin et al., "Microsoft COCO: Common objects in context," in Eur. Conf. Comput. Vis., 2014, pp. 740-755. DOI: 10.1007/978-3-319-10602-1_48.

[33] M. Anthimopoulos et al., "Computer vision-based carbohydrate estimation for type 1 patients with diabetes using smartphones," J. Diabetes Sci. Technol., vol. 8, pp. 1444-1449, 2014. DOI: 10.1177/1932296814528945.

[34] Y. Kawano and K. Yanai, "Automatic expansion of a food image dataset leveraging existing categories with domain adaptation," in Eur. Conf. Comput. Vis. Workshops, 2014, pp. 3-17. DOI: 10.1007/978-3-319-16199-0_1.

[35] F. Zhu et al., "The use of mobile devices in aiding dietary assessment and evaluation," IEEE J. Sel. Top. Signal Process., vol. 4, pp. 756-766, 2010. DOI: 10.1109/JSTSP.2010.2051471.

[36] I. S. Martín et al., "Food recognition and nutritional information system," IEEE Access, vol. 8, pp. 77036-77048, 2020. DOI: 10.1109/ACCESS.2020.2986000.

[37] A. Salvador et al., "Learning cross-modal embeddings for cooking recipes and food images," in IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 3020-3028. DOI: 10.1109/CVPR.2017.327.

[38] J. Marin et al., "Recipe1M+: A dataset for learning cross-modal embeddings for cooking recipes and food images," IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, pp. 187-203, 2019. DOI: 10.1109/TPAMI.2019.2927476.

[39] J. Chen et al., "Cross-modal recipe retrieval: How to cook this dish?" in Int. Conf. Multimedia Model., 2017, pp. 588-600. DOI: 10.1007/978-3-319-51811-4_9.

[40] J. Yosinski et al., "How transferable are features in deep neural networks?" in Adv. Neural Inf. Process. Syst., 2014. DOI: 10.48550/arXiv.1411.1792.

[41] S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Trans. Knowl. Data Eng., vol. 22, pp. 1345-1359, 2010. DOI: 10.1109/TKDE.2009.191.

[42] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," J. Big Data, vol. 6, no. 60, 2019. DOI: 10.1186/s40537-019-0197-0.

[43] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," arXiv preprint arXiv:1712.04621, 2017. DOI: 10.48550/arXiv.1712.04621.

[44] D. M. W. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," J. Mach. Learn. Technol., vol. 2, pp. 37-63, 2011. DOI: 10.48550/arXiv.2010.16061.

[45] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," Inf. Process. Manag., vol. 45, pp. 427-437, 2009. DOI: 10.1016/j.ipm.2009.03.002.