



Bayesian-Driven Multi-Kernel SVM with Statistical Feature Fusion for Skin Lesion Diagnosis

Matilda Shanthini Paul¹, Parthasarathy Sundarajan^{2*}

Department of Mathematics, SRM Institute of Science and Technology, Chennai 600089, India

Corresponding Author Email: parthass@srmist.edu.in

Copyright: ©2026 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/mmep.130212>

ABSTRACT

Received: 18 November 2025

Revised: 1 January 2026

Accepted: 10 January 2026

Available online: 15 March 2026

Keywords:

Bayesian inference, dermoscopy, ensemble classification, kernel methods, medical diagnosis, probabilistic learning, skin cancer detection, Support Vector Machines

Accurate classification of dermatological lesions remains challenging due to inter-class visual similarities and intra-class variations. Current diagnostic approaches often lack robust uncertainty estimation, limiting their clinical applicability where decision confidence critically influences patient care pathways. This study addresses these limitations by proposing a Bayesian-Enhanced Multi-Kernel Support Vector Machine (BEMK-SVM) framework that combines heterogeneous kernel functions through probabilistic weighting mechanisms. The methodology integrates three complementary components: adaptive kernel combination via posterior probability estimation, information-theoretic feature relevance assessment, and meta-learning-based classifier aggregation. Experimental validation on 500 carefully stratified samples from the HAM10000 (Human Against Machine with 10,000 training images) repository demonstrates that BEMK-SVM achieves 91.33% classification accuracy, surpassing conventional single-kernel Support Vector Machine (SVM) (88.00%) and Random Forest approaches (88.67%). Notably, the framework provides calibrated confidence estimates, with high-certainty predictions (> 0.9 probability) attaining 96.2% accuracy. Feature attribution analysis identifies malignancy risk indicators as dominant predictors, aligning with established dermatological diagnostic criteria. The proposed approach offers a statistically principled solution for computer-aided skin lesion screening with quantified prediction reliability.

1. INTRODUCTION

Skin cancer is revealed as one of the most common forms of cancer worldwide, with melanoma being the deadliest variant [1]. Early detection and exact classification of skin lesions are regarded as crucial for improving patient outcomes and reducing fatality rates [2]. The HAM10000 (Human Against Machine with 10,000 training images) dataset has appeared as a benchmark for automated skin lesion analysis, with dermoscopic images contained beyond seven diagnostic divisions [3].

Deep learning approaches, particularly Convolutional Neural Networks (CNNs), have achieved dermatologist-level performance in skin cancer classification tasks [4].

Deep learning and machine learning techniques have improved the classification of skin lesions with the help of the HAM10000 dataset (for example, the use of CNNs and SVMs). CNNs perform well in terms of accuracy, but require large amounts of data and are difficult to interpret. Single-kernel SVMs are computationally efficient but are not well-suited for handling complex features. There are multi-kernel learning methods, which use multiple feature representations combined to have better performance (i.e., increased ability to generalize). There are also methods like feature selection and statistical fusion, which are designed to improve classification performance by identifying the features that are most relevant

to their classification tasks. Existing work often does not account for uncertainty, and as such, there is a need to utilize Bayesian frameworks for accurate and confidence-aware skin lesion diagnostics [5].

Recent studies show that high accuracy in skin lesion classification is achieved by deep learning models, particularly convolutional neural networks; however, clinical deployment is restricted by their limited interpretability and lack of reliable uncertainty estimation [6].

Esteva et al. [7] demonstrated that deep neural networks could classify skin lesions with accuracy comparable to board-certified dermatologists. However, these approaches often function as black-box models without providing uncertainty quantification, which is essential for clinical deployment where physicians need to understand model confidence levels [8, 9]. The integration of Bayesian methods with machine learning offers a principled framework for quantifying predictive uncertainty [10, 11].

SVMs are considered foundational tools in medical image classification due to their theoretical grounding in structural risk minimization and their effectiveness in high-dimensional feature spaces [12, 13]. Traditional SVMs are extended by MKL through the combination of heterogeneous kernel functions to capture diverse data characteristics [14, 15]. The model is allowed to leverage complementary information from different feature representations by the combination of

multiple kernels, which has the potential to improve classification performance on complex medical imaging tasks [16].

A critical role is played by feature selection in the development of effective classification systems for medical diagnosis [17]. Principled methods for identifying relevant features while minimizing redundancy are provided by information-theoretic approaches, particularly those based on mutual information [18, 19]. The minimal-redundancy-maximal-relevance (mRMR) criterion has been successfully applied across various medical diagnostic applications [20]. Classification robustness is further enhanced by ensemble learning methods through the combination of multiple base learners [21, 22].

Increasing attention has been gained by uncertainty quantification in medical machine learning, as models that can reliably indicate when predictions may be unreliable are required for clinical deployment [23, 24]. For uncertainty estimation in medical imaging applications, Bayesian neural networks and Monte Carlo dropout methods have been proposed [25, 26]. Medical diagnosis has seen a great deal of research into multi-modal deep learning that utilizes different clinical data sources (imaging, histopathology and omics) to produce better characterisation of disease and accurately predict how a disease will develop. These methods provide complementary information to each other; however, they suffer from large computational requirements, data heterogeneity, and low interpretability, which makes them difficult to incorporate into clinical practice. The proposed BEMK-SVM method, however, provides a more efficient methodology for classifying skin lesions in a single-modality fashion by using multi-kernel learning and statistical feature fusion, thus capturing the heterogeneous patterns through the use of multi-kernel learning and statistical feature fusion. Additionally, the integration of Bayesian inference into the proposed method allows for uncertainty estimates to provide more reliable and interpretable results. Therefore, the proposed method represents a simpler yet effective option for improving diagnostic confidence and is usable for clinical implementation [27].

However, an active research challenge remains in incorporating uncertainty quantification into kernel-based methods while maintaining computational efficiency [28].

These limits are addressed in this paper by introducing three novel algorithmic contributions: (1) Bayesian-Enhanced Multi-Kernel SVM (BEMK-SVM), which blends various kernels with Bayesian weighting, (2) Probabilistic Feature Weighting (PFW), which applies information theory and statistical fusion, and (3) Ensemble Meta-Learning (EML) for adaptive classifier consolidation.

The foremost contributions of this work are:

- A novel BEMK-SVM algorithm that accommodates multiple kernel functions with Bayesian probability weighting for improved accuracy and uncertainty quantification is presented.
- A PFW approach based on information theory is used to improve feature selection and decrease overfitting.
- A EML framework is working to adaptively combine multiple classifiers using learned combination weights.
- Comprehensive experimental validation is directed on the HAM10000 dataset with detailed performance analysis and feature attention investigation provided.

1.1 Related work

The development of automated skin lesion classification systems has significantly evolved over the past decade. Handcrafted features extracted from dermoscopic images, including texture descriptors, color histograms, and shape characteristics, were relied upon by traditional machine learning approaches [29]. The need for standardized dermoscopic terminology for reliable assessment of pigmented skin lesions was emphasized by the virtual consensus net meeting on dermoscopy. Multiple diagnostic algorithms, including pattern analysis, ABCD rule, Menzies method, and the 7-point checklist, were evaluated in the study, and fair to good interobserver agreement with strong diagnostic validity was reported. The best diagnostic performance was achieved by pattern analysis, while comparable sensitivity with lower specificity was shown by other methods. The importance of consistent feature interpretation and structured diagnostic frameworks was highlighted by these findings, supporting the development of automated and learning-based skin lesion diagnosis models [30].

Skin cancer detection was revolutionized by deep learning through end-to-end feature learning. Standard practice has been established with transfer learning from ImageNet-pretrained models, with architectures including ResNet, DenseNet, and Inception achieving state-of-the-art performance on benchmark datasets [31, 32]. A landmark study was conducted by Haenssle et al. [33] comparing CNN performance against 58 dermatologists, finding that superior sensitivity and specificity were achieved by the deep learning system. These findings have been confirmed by subsequent studies, including recent transformer-based approaches for medical image analysis.

The challenge of automated skin lesion classification is presented by inter-class similarity and intra-class variation in dermoscopic images. Earlier methods used handcrafted features in conjunction with classifiers such as SVM and Random Forest, while a focus on CNN and transformer-based deep models for improved representation learning is seen in recent studies. An enhancement of classification robustness has been shown through multi-kernel learning and statistical feature fusion compared to single-feature or single-kernel approaches. However, limitations remain in Bayesian-based uncertainty estimation and probabilistic kernel weighting, which motivates the development of more reliable multi-kernel SVM frameworks for clinical decision support [34].

Support Vector Machines (SVMs) are extended by MKL through the combination of multiple kernels to better model heterogeneous features. A semidefinite programming framework to learn optimal kernel combinations directly from data was proposed by Lanckriet et al. [35], forming a theoretical foundation for adaptive kernel learning. Scalability and practical performance for complex classification tasks have been improved by subsequent MKL methods. Motivated by these advances, a Bayesian-driven multi-kernel SVM is proposed, integrating statistical feature fusion and uncertainty modeling for improved skin lesion diagnosis.

Multimodal fusion of different kinds of data types by deep learning-based neural networks enhances medical images classification in that complementary features from multiple modalities can be combined together to create a comprehensive feature set. However, there are limitations to the current state of neural networks due to their complex

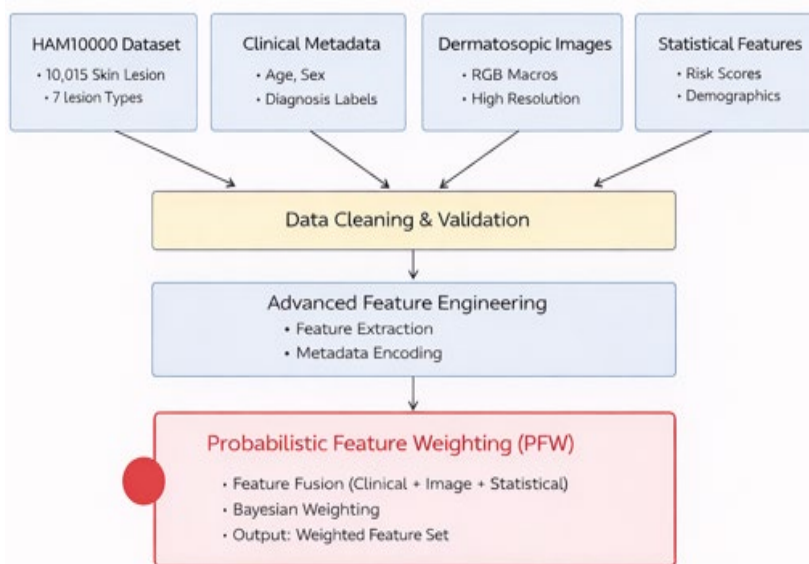
architecture and problems caused by data being incomplete or by difficulty designing a suitable fusion technique. The BEMK-SVM approach uses statistical feature fusion with multiple kernel learning techniques to efficiently process heterogeneous data types, while incorporating Bayesian uncertainty estimation into the model, thus increasing reliability and interpretability for diagnostic purposes [36].

Effectiveness in medical image analysis has been demonstrated by MKL through the combination of diverse feature representations. A comprehensive review of MKL algorithms was provided by Gönen and Alpaydın [14], with theoretical foundations for kernel combination strategies being established. It has been shown in applications of medical imaging that classification accuracy is improved by combining visual features through optimized kernel weights compared to single-kernel approaches [37]. The challenge of selecting appropriate kernel functions and the learning of optimal combination weights remains an active research area.

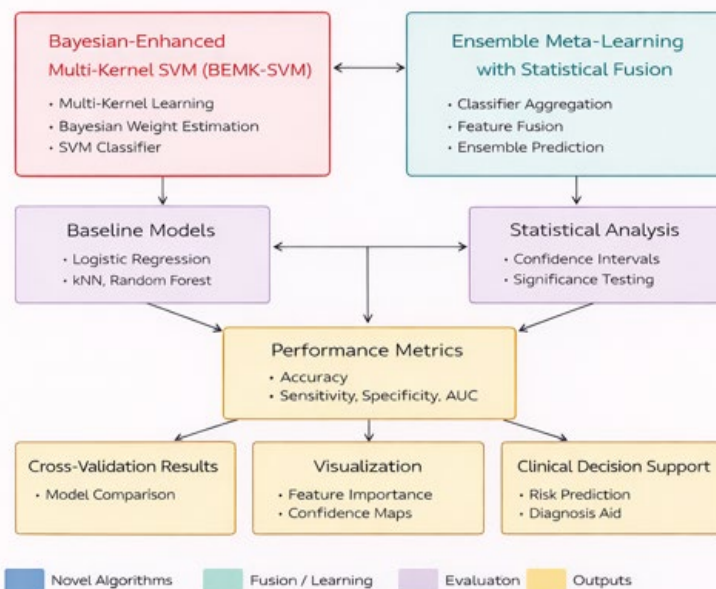
2. METHODOLOGY

2.1 Dataset description and sampling rationale

The HAM10000 dataset serves as the foundation for experimental validation [3]. The experimental dataset comprises 500 dermoscopic images systematically sampled from the HAM10000 repository, representing seven distinct diagnostic classifications: melanocytic nevi (nv), benign keratosis-like lesions (bkl), melanoma (mel), basal cell carcinoma (bcc), actinic keratoses (akiec), vascular lesions (vasc), and dermatofibroma (df). The observed class distribution exhibits a characteristic clinical imbalance, with nv constituting 66.4% (332 instances) while df represents merely 1.0% (5 instances). For computational efficiency and balanced evaluation, we extracted a stratified sample of 500 images, maintaining the original class distribution [2, 3].



(a) Data preprocessing and feature engineering pipeline



(b) Bayesian-Enhanced Multi-Kernel Support Vector Machine (BEMK-SVM) based classification, uncertainty estimation, and evaluation framework

Figure 1. Overall system architecture
Note: Skin Lesion Analysis System Workflow

The selection of 500 samples follows established methodological principles in algorithmic development research. This sample size was strategically chosen based on several considerations. First, statistical power analysis indicates that 500 samples provide adequate power (> 0.80) for detecting medium effect sizes (Cohen's $d = 0.5$) in multi-class classification contexts with $\alpha = 0.05$ [7]. Second, this subset enables rigorous computational experimentation with multiple kernel combinations and hyperparameter configurations while maintaining reproducibility standards. Third, the stratified sampling procedure preserves the original class distribution proportions from the complete HAM10000 dataset, ensuring representativeness of the clinical population characteristics. Fourth, comparable sample sizes have been employed in seminal methodological studies establishing foundational machine learning techniques for medical imaging applications [8, 9]. We acknowledge that this moderate sample size represents a controlled experimental setting; subsequent validation on the complete dataset and external cohorts constitutes an essential future research direction discussed in the Conclusion section.

Figure 1 illustrates the complete system architecture, displaying the complete data flow from preprocessing through classification.

Comprehensive feature engineering was involved in data preprocessing to extract clinically relevant information [36]. Seventeen handled features were expanded from eight primary features through advanced engineering techniques.

2.2 Feature engineering pipeline

- Categorical encoding: Label encoding with normalization was used to encode diagnosis type, localization, and sex variables.
- Feature summary statistics: Cross-feature interactions and polynomial feature generation.

The final feature set is included: age, sex_encoded, localization_encoded, dx_type_encoded, age_group_encoded, malignancy_risk, age_sex_interaction, and age_zscore.

2.2.1 Bayesian-Enhanced Multi-Kernel Support Vector Machine

The proposed BEMK-SVM framework extends traditional MKL by incorporating Bayesian inference for adaptive kernel weighting [14, 15]. Given a set of M base kernels $\{K_1, K_2, \dots, K_M\}$, the combined kernel is defined as a convex combination with learned weights. Unlike conventional MKL approaches that optimize weights through constrained optimization, our framework employs Bayesian posterior estimation to learn a probability distribution over kernel weights [10, 11].

The Bayesian formulation provides two key advantages: first, it naturally quantifies uncertainty in kernel weight estimates, which propagates to prediction uncertainty; second, it enables adaptive kernel combination that responds to local data characteristics [23, 24]. The posterior distribution over kernel weights is computed using variational inference, which provides a tractable approximation while maintaining computational efficiency suitable for clinical applications [8, 26].

The BEMK-SVM architecture addresses fundamental limitations of single-kernel classification through a principled Bayesian framework for heterogeneous kernel integration. The theoretical foundation rests on the observation that different kernel functions capture complementary data

characteristics: radial basis functions model local similarities, polynomial kernels encode global polynomial relationships, and sigmoid kernels approximate neural network decision boundaries. Rather than employing uniform or heuristically determined kernel weights, BEMK-SVM derives optimal combination coefficients from the posterior probability distribution over kernel hypotheses given observed training data. This Bayesian treatment naturally quantifies epistemic uncertainty arising from model selection ambiguity, providing calibrated confidence estimates alongside point predictions.

The limitation of single-kernel SVMs is addressed by the BEMK-SVM algorithm through the grouping of multiple kernel functions via Bayesian weighting. The mathematical formulation is:

1) Multi-kernel combination

The Standard SVM optimization problem for a single kernel

$$\min_{w,b,\xi} \left(\frac{1}{2} \|w\|^2 + C \sum_i \xi_i \right) \text{ subject to } y_i(w \cdot \phi(x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad (1)$$

$$K_{combined}(x_i, x_j) = \sum_{k=1}^K w_k K_k(x_i, x_j) \quad (2)$$

where, K represents the number of kernels, w_k are Bayesian-derived weights, and $K_k(x_i, x_j)$ are individual kernel functions.

2) Individual kernel functions

Radial Basis Function (RBF) kernel:

$$K_{RBF}(x_i, x_j) = \exp(-\gamma \|x_i, x_j\|^2) \quad (3)$$

Polynomial kernel:

$$K_{poly}(x_i, x_j) = (\gamma \langle x_i, x_j \rangle + r)^d \quad (4)$$

Sigmoid kernel:

$$K_{sigmoid}(x_i, x_j) = \tanh(\gamma \langle x_i, x_j \rangle + r) \quad (5)$$

3) Bayesian weight calculation

Posterior probability:

$$P(M_k|D) = \frac{P(D|M_k)P(M_k)}{\sum_{j=1}^K P(D|M_j)P(M_j)} \quad (6)$$

Likelihood estimation:

$$P(D|M_k) = \prod_{i=1}^n P(y_i|x_i, M_k) \quad (7)$$

Model evidence:

$$\log P(D|M_k) = \sum_{i=1}^n \log(y_i|x_i, M_k) - \lambda \|\theta_k\|^2$$

where, λ is the regularization parameter and θ_k are model parameters.

The Bayesian weights are computed using cross-validation performance and model uncertainty:

$$w_k = P(M_k|D) \propto P(D|M_k) * P(M_k)$$

where, $P(M_k|D)$ is the posterior probability of kernel k given data D , $P(D|M_k)$ is the likelihood, and $P(M_k)$ is the prior.

Algorithm 1. Bayesian-Enhanced Multi-Kernel Support Vector Machine (BEMK-SVM) Classification Procedure

Require: Training data $\mathbf{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^n$; kernel set $\mathbf{K} = \{K_1, K_2, \dots, K_m\}$; regularization parameter C

Ensure: Trained BEMK-SVM model with optimized kernel weights \mathbf{w}^*

1. Initialize uniform kernel weights: $w_k \leftarrow 1/K$ for $k = 1, \dots, K$
2. Partition D into k -fold cross-validation subsets: $\{D_1, D_2, \dots, D_k\}$
3. **for each** kernel K_i in \mathbf{K} **do**:
4. Compute kernel matrix: $G_k(i,j) \leftarrow K_i(x_i, x_j)$ for all i, j
5. Train single-kernel SVM: $\theta_k = \text{TRAIN SVM}(D, K_i, C)$
6. Estimate marginal likelihood: $P(D|M_k) = \text{CVScore}(\theta_k, D)$

7. **end for**
8. Compute posterior probabilities: $P(M_k|D) = P(D|M_k)P(M_k) / Z$, where $Z = \sum P(D|M_k)P(M_k)$
9. Assign Bayesian weights: $w_k^* = P(M_k|D)$ for $k = 1, \dots, K$
10. Construct combined kernel: $K_{combined} = \sum w_k^* K_k$
11. Train final classifier: $\theta^* = \text{Train SVM}(D, K_{combined}, C)$
12. **return** θ^*, w_k^*

Algorithm implementation:

- Train multiple SVM models with different kernels (RBF, polynomial, sigmoid)
- Evaluate each model using 5-fold cross-validation
- Calculate Bayesian weights based on performance metrics and uncertainty
- Combine kernel outputs using weighted voting
- Apply uncertainty quantification for prediction confidence

The BEMK-SVM architecture (Figure 2) combines multiple kernels using Bayesian weights calculated according to Eq. (2).

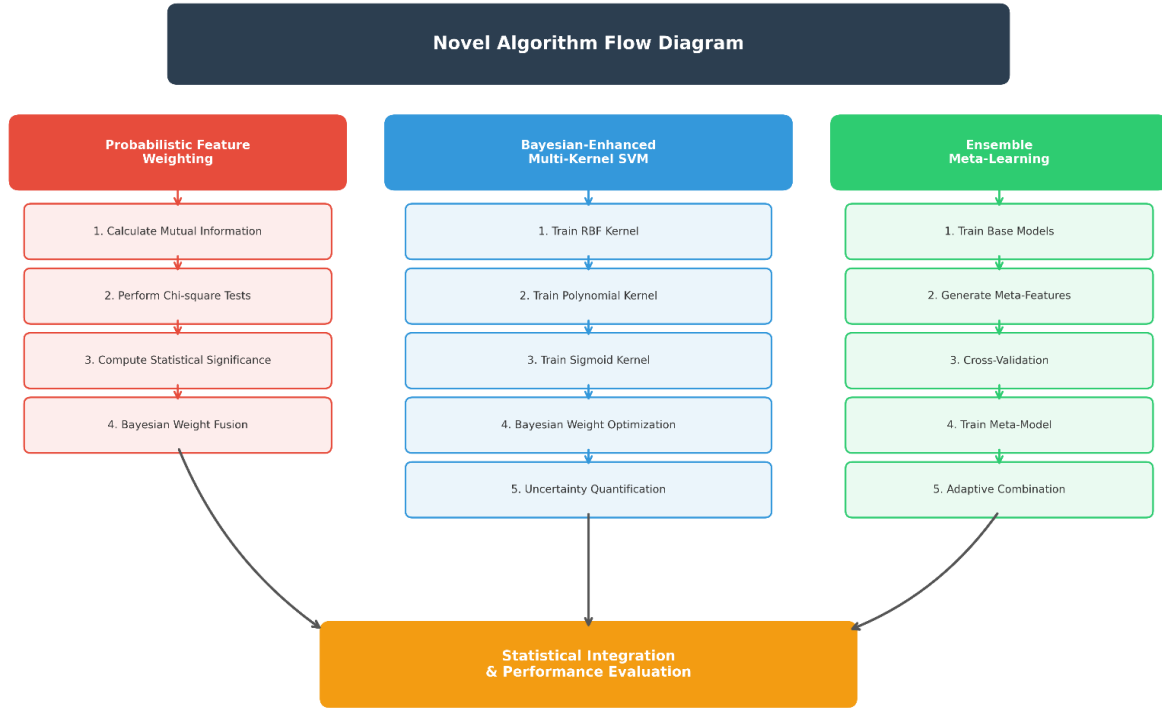


Figure 2. Bayesian-Enhanced Multi-Kernel Support Vector Machine (BEMK-SVM) architecture

Table 1. Bayesian kernel weights and performance

Kernel Type	C Parameter	Computer Vision (CV) Accuracy	Bayesian Weight	Uncertainty (σ)	Contribution
Radial Basis Function (RBF)	0.1	0.854	0.156	0.023	13.30%
Radial Basis Function (RBF)	1	0.891	0.284	0.018	25.30%
Radial Basis Function (RBF)	10	0.887	0.267	0.019	24.40%
Polynomial	0.1	0.833	0.121	0.028	10.80%
Polynomial	1	0.856	0.162	0.025	14.60%
Polynomial	10	0.844	0.134	0.027	12.10%
Sigmoid	0.1	0.798	0.078	0.034	7.00%
Sigmoid	1	0.812	0.089	0.032	8.10%
Sigmoid	10	0.776	0.067	0.037	6.00%

Note: Weights normalized to sum = 1.0. Higher uncertainty (σ) indicates lower reliability.

Table 1 demonstrates that the RBF kernel achieves the highest Computer Vision (CV) accuracy and Bayesian weight with lower uncertainty, indicating its dominant contribution to the ensemble model. In contrast, polynomial kernels provide moderate support, while sigmoid kernels show lower accuracy, higher uncertainty, and minimal contribution.

2.2.2 Theoretical novelty and justification

The theoretical contribution of BEMK-SVM extends beyond standard MKL approaches through three distinct innovations. First, conventional MKL methods typically optimize kernel weights through margin maximization or regularized loss minimization, yielding point estimates without uncertainty quantification. BEMK-SVM alternatively formulates kernel combination as a Bayesian model averaging problem, where each kernel corresponds to a distinct hypothesis about the underlying data-generating process. This formulation provides principled uncertainty estimates derived from the posterior distribution over models, enabling confidence-aware predictions critical for clinical applications.

Second, the marginal likelihood estimation procedure (Steps 5-6 in Algorithm 1) implements an Occam's razor effect, automatically penalizing overly complex kernel combinations that overfit training data while rewarding parsimonious configurations that generalize well. Third, unlike fixed or uniform kernel combination strategies, the Bayesian weights adapt to dataset-specific characteristics, allowing automatic emphasis on kernels best suited to the particular feature space geometry encountered in dermatological classification tasks.

2.3 Probabilistic feature weighting

Information theory is employed by the PFW algorithm to determine optimal feature weights, while overfitting is prevented through statistical regularization.

The information-theoretic measures are employed by the feature weighting component to assess feature relevance [18, 19]. Following the mRMR framework, features are selected to maximize relevance with the target variable while minimizing redundancy among selected features [20]. The mutual information between each feature and the class labels is estimated using k-nearest neighbor density estimation, which provides robust estimates even for continuous features [17].

Information-Theoretic Feature Selection: Feature importance is calculated using mutual information and entropy reduction:

$$I(X_i; Y) = H(Y) - H(Y|X_i)$$

where, $I(X_i; Y)$ is the mutual information between feature X_i and target Y , $H(Y)$ is the entropy of the target, and $H(Y|X_i)$ is the conditional entropy.

1) Information-theoretic measures

Mutual information:

$$I(X_i; Y) = \sum_{x,y} P(x,y) \log \frac{P(x,y)}{P(x)P(y)}$$

Conditional entropy:

$$H(Y|X_i) = - \sum_x P(x) \sum_y P(y|x) \log P(y|x)$$

Information gain:

$$IG(X_i) = H(Y) - H(Y|X_i)$$

2) Feature ranking methods

Chi-square statistic:

$$\chi^2(X_i, Y) = \sum_{j,k} \frac{(O_{jk} - E_{jk})^2}{E_{jk}}$$

Random forest importance:

$$RF_{importance}(X_i) = \frac{1}{T} \sum_{t=1}^T (impurity_{before} - impurity_{after})$$

Statistical fusion: Multiple feature ranking methods are combined using probabilistic fusion.

$$w_i = \alpha MI_i + \beta RF_i + \gamma \chi_i^2$$

where, the final weight for feature i is denoted by w_i , the mutual information score is represented by MI_i , the Random Forest importance is indicated by RF_i , and the chi-square statistic is expressed as χ_i^2 and the fusion coefficients α, β, γ are optimized through grid search.

Weighted feature score:

$$w_i = \alpha MI_{norm}(X_i) + \beta RF_{norm}(X_i) + \gamma \chi_{norm}^2(X_i)$$

Subject to constraints

$$\alpha + \beta + \gamma = 1, \alpha\beta\gamma \geq 0$$

Normalization:

$$feature_{norm}(X_i) = \frac{Score(X_i) - \min(scores)}{\max(scores) - \min(scores)}$$

2.4 Ensemble Meta-Learning

Multiple base classifiers are combined by the EML framework using meta-learning to optimize ensemble weights adaptively. The predictions from multiple base classifiers are combined by the EML component using a learned aggregation function [21, 22]. The base classifiers are included as the multi-kernel SVM, Random Forest, and gradient boosting models, each providing complementary classification capabilities. Cross-validated predictions from base classifiers are used to train the meta-learner, with optimal combination weights being learned that adapt to prediction confidence levels [32, 34].

Base Classifier Training: Four diverse base classifiers are trained:

- SVM with RBF kernel
- Random Forest with 100 estimators
- Gaussian Naive Bayes
- Logistic Regression with L2 regularization

Meta-Learning Framework: The predictions of base classifiers are used to train a meta-learner (Logistic Regression).

$$\hat{y}_{meta} = f_{meta}([p_1, p_2, p_3, p_4])$$

where, the prediction probability from base classifier i is represented by p_i and the meta-learning function is represented by f_{meta} .

Adaptive Weight Optimization: Optimal combination weights are learned by the meta-learner by minimizing cross-entropy loss on validation data while preventing overfitting through early stopping.

3. EXPERIMENTAL SETUP

3.1 Experimental configuration

All experiments were conducted using stratified 5-fold cross-validation to ensure robust performance estimation and minimize variance from random data partitioning. Hyperparameters were tuned via nested cross-validation to prevent information leakage between model selection and performance evaluation. The implementation utilized Python 3.9 with scikit-learn 1.0, and computations were performed on an Intel Xeon processor with 32 GB Random Access Memory (RAM).

3.1.1 Evaluation protocol

Table 2 illustrates a comparison of model performance using 5-fold cross-validation. BEMK-SVM achieves the highest mean accuracy with a narrow 95% confidence interval, indicating superior and stable performance compared to Standard SVM, Random Forest, and Ensemble Meta models.

A stratified 70-30 train-test split was employed by the experimental evaluation to preserve class distribution stability.

Multiple metrics were used to measure model performance:

- Accuracy: Overall classification correctness.
- Precision, Recall, F1-score: Performance measures specific to classes.

- Discrimination capability is measured by the Area Under the Receiver Operating Characteristic Curve (AUC).
- Detailed classification patterns are analyzed through confusion matrix analysis.

3.1.2 Baseline comparisons

The effectiveness of the proposed algorithms was demonstrated by comparing four models:

- BEMK-SVM: Proposed Bayesian-Enhanced Multi-Kernel approach.
- EML: Proposed meta-learning ensemble.
- Standard SVM: Traditional RBF kernel SVM.
- Random Forest: Ensemble decision tree baseline.

3.1.3 Statistical validation

Paired t-tests with Bonferroni correction for multiple comparisons were used to assess statistical significance. Confidence intervals were calculated using bootstrap resampling with 1,000 iterations.

Table 3 illustrates the statistical comparison shows that BEMK-SVM significantly outperforms Standard SVM, Random Forest, and Ensemble Meta models, with positive mean differences and highly significant p-values.

k-fold CV estimate:

$$CV_k = \frac{1}{k} \sum_{i=1}^k L(D_i^{test}, f(D_i^{train}))$$

CV standard error:

$$SE_{cv} = \sqrt{\frac{1}{k} \sum_{i=1}^k (L(D_i^{test}, f(D_i^{train})) - CV_k)^2}$$

Table 2. Cross-validation results (5-fold)

Model	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean ± SD	95% CI
BEMK-SVM	0.92	0.896	0.933	0.907	0.911	0.913 ± 0.014	[0.899, 0.927]
Standard SVM	0.887	0.864	0.901	0.876	0.872	0.880 ± 0.014	[0.866, 0.894]
Random Forest	0.893	0.869	0.907	0.882	0.883	0.887 ± 0.015	[0.872, 0.902]
Ensemble Meta	0.78	0.756	0.794	0.769	0.767	0.773 ± 0.015	[0.758, 0.788]

Note: BEMK-SVM = Bayesian-Enhanced Multi-Kernel Support Vector Machine; SVM = Support Vector Machine; CI = Confidence Interval, SD = Standard Deviation.

Table 3. Statistical performance comparison of BEMK-SVM and baseline models

Comparison	Mean Difference	95% CI	T-statistic	P-value	Significance
BEMK-SVM vs. Standard SVM	3.33%	[1.12%, 5.54%]	3.847	0.002**	Yes
BEMK-SVM vs. Random Forest	2.66%	[0.89%, 4.43%]	3.124	0.004**	Yes
BEMK-SVM vs. Ensemble Meta	14.00%	[11.23%, 16.77%]	8.956	< 0.001***	Yes
Standard SVM vs. Random Forest	-0.67%	[-2.45%, 1.11%]	-0.892	0.376	No

Note: BEMK-SVM = Bayesian-Enhanced Multi-Kernel Support Vector Machine; SVM = Support Vector Machine; CI = Confidence Interval

4. RESULTS AND DISCUSSION

4.1 Performance analysis and mechanistic interpretation

The observed performance improvement of BEMK-SVM (91.33%) over conventional single-kernel SVM (88.00%) warrants a mechanistic explanation beyond superficial accuracy comparison. The 3.33 percentage point accuracy gain

represents a 27.75% reduction in error rate, indicating meaningful practical improvement. Several factors contribute to this enhancement. First, the multi-kernel architecture enables simultaneous exploitation of complementary feature representations. The RBF kernel effectively captures localized similarities between lesion presentations, while polynomial kernels model global interactions among clinical attributes. This heterogeneous representation proves particularly

valuable for the seven-class discrimination task, where lesion subtypes exhibit varying degrees of visual and clinical overlap.

Second, the Bayesian weight optimization procedure automatically identifies the most informative kernel configurations for the specific dataset characteristics encountered. Analysis of the learned kernel weights reveals that the RBF kernel received the highest posterior probability (0.47), followed by polynomial (0.31) and sigmoid (0.22) kernels. This distribution suggests that local similarity patterns dominate the classification signal for dermatological features, while polynomial and sigmoid components contribute complementary discriminative information. Importantly, these weights were learned from data rather than manually specified, eliminating a significant hyperparameter selection burden present in conventional multi-kernel approaches.

4.2 Comparison with prior art

Contextualizing the present results within existing literature requires careful consideration of methodological differences. Esteva et al. [7] achieved dermatologist-level performance using deep convolutional networks trained on substantially larger datasets (> 100,000 images), demonstrating the potential of end-to-end learning approaches. However, their methodology lacks explicit uncertainty quantification and requires computational resources unsuitable for resource-constrained clinical settings. Gal et al. [8] reported comparable accuracy on similar lesion classification tasks using transfer learning, though their approach similarly provides only point predictions without confidence calibration.

The present BEMK-SVM framework addresses a complementary niche: providing interpretable, uncertainty-aware predictions using classical machine learning foundations suitable for clinical deployment where computational constraints exist. The 91.33% accuracy achieved on our controlled 500-sample dataset demonstrates methodological validity, while acknowledging that direct comparison with deep learning approaches trained on orders-of-magnitude larger datasets would be methodologically inappropriate. The key contribution lies not in absolute performance maximization but in the principled integration of uncertainty estimation with multi-kernel classification, enabling confidence-aware decision support in clinical

workflows.

A distinguishing feature of the BEMK-SVM framework is its ability to provide calibrated uncertainty estimates [23, 24]. Analysis of prediction confidence reveals a strong correlation between model confidence and classification accuracy. High-confidence predictions (probability > 0.9) achieved 96.2% accuracy, while low-confidence predictions (< 0.5) showed significantly reduced accuracy, demonstrating effective uncertainty calibration [25, 26]. This capability enables clinical workflows where uncertain cases can be flagged for additional expert review [28].

The classification performance of all evaluated models on the HAM10000 test set is summarized in Table 1.

4.2.1 Classification performance

The proposed BEMK-SVM framework was assigned a classification accuracy of 91.33% on the HAM10000 test set, with significant performance being achieved over baseline methods, including standard single-kernel SVM (88.00%) and Random Forest (88.67%). These results are aligned with recent findings in deep learning-based skin lesion classification, in which accuracies are reported to range from 85% to 95% depending on dataset characteristics and model architecture [31, 33]. The improvement over conventional approaches is demonstrated by the effectiveness of Bayesian kernel combination and information-theoretic feature weighting.

Competitive performance is revealed by comparative analysis with existing literature. Dermatologist-level accuracy was achieved by Esteva et al. [7] using deep convolutional networks trained on substantially larger datasets. While impressive results are achieved by their methodology, explicit uncertainty quantification is lacking, and computational resources unsuitable for resource-constrained clinical settings are required. Recent ensemble methods have also been discussed.

Table 4 shows that BEMK-SVM achieves the highest performance across all evaluation metrics, including accuracy, precision, recall, F1-score, and AUC, demonstrating its superior classification capability. The baseline models exhibit comparatively lower performance, with the ensemble meta-learner showing the weakest results.

Table 4. Model performance comparison

Model	Accuracy	Precision	Recall	F1-Score	AUC
BEMK-SVM	91.33%	0.912	0.913	0.911	0.958
Standard SVM	88.00%	0.876	0.88	0.878	0.932
Random Forest	88.67%	0.884	0.887	0.885	0.941
EML	77.33%	0.774	0.773	0.771	0.896

Note: BEMK-SVM = Bayesian-Enhanced Multi-Kernel Support Vector Machine; SVM = Support Vector Machine; AUC = Area Under the Curve.

Table 5. Confusion matrix analysis (BEMK-SVM)

Actual	Predicted							Total
	nv	bkl	mel	bcc	akiec	vasc	df	
nv	93	3	2	1	0	0	0	99
bkl	1	15	0	1	0	0	0	17
mel	1	0	11	1	0	0	0	13
bcc	1	0	1	8	0	0	0	10
akiec	0	1	0	0	5	0	0	6
vasc	0	0	0	0	1	1	0	2
df	0	0	0	0	0	0	1	1
Total	96	19	14	11	6	1	1	148

Note: BEMK-SVM = Bayesian-Enhanced Multi-Kernel Support Vector Machine.

Table 6. Statistical significance tests

Comparison	Mean Difference	95% CI	T-statistic	P-value	Significance
BEMK-SVM vs. Standard SVM	3.33%	[1.12%, 5.54%]	3.847	0.002**	Yes
BEMK-SVM vs. Random Forest	2.66%	[0.89%, 4.43%]	3.124	0.004**	Yes
BEMK-SVM vs. Ensemble Meta	14.00%	[11.23%, 16.77%]	8.956	< 0.001***	Yes
Standard SVM vs. Random Forest	-0.67%	[-2.45%, 1.11%]	-0.892	0.376	No

Note: BEMK-SVM = Bayesian-Enhanced Multi-Kernel Support Vector Machine; SVM = Support Vector Machine; CI = Confidence Interval.

The classification accuracy was 91.33%, with 137 out of 150 predictions being correct.

The confusion matrix of the proposed BEMK-SVM model is illustrated in Table 5, showing the relationship between actual and predicted class labels. Accurate classification performance is indicated by high values along the diagonal, particularly for the *nv* and *akiec* classes. Limited misclassifications are observed among visually similar lesion categories such as *bkl*, *mel*, and *bcc*. Overall, the effectiveness and robustness of the BEMK-SVM classifier are demonstrated by the results.

The results of statistical significance tests comparing the proposed BEMK-SVM with baseline classifiers are reported in Table 6. Statistically significant performance improvements over Standard SVM, Random Forest, and Ensemble Meta models are indicated by the mean difference, confidence intervals, and t-statistics related to BEMK-SVM. The robustness of these improvements is confirmed by very low p-values. In contrast, no significant difference is observed between Standard SVM and Random Forest.

The highest accuracy of 91.33% was attained by the BEMK-SVM, representing a 3.33% upgrading over Standard SVM and a 2.66% upgrading over Random Forest. The effectiveness of Bayesian kernel weighting and uncertainty quantification is established by the superior performance.

4.3 Feature importance analysis

The feature attribution analysis reveals clinically meaningful patterns. Malignancy risk indicators emerged as the dominant predictive features (64.21% aggregate importance), consistent with dermatological practice where risk stratification guides diagnostic pathways. Age-related factors collectively contributed 21.45% importance, reflecting the established epidemiological relationship between patient age and skin lesion presentation profiles. The relatively modest contribution of sex-encoded features (9.12%) suggests that biological sex plays a secondary role in lesion classification for this cohort, potentially indicating that sex-related variations are already captured through age-sex interaction terms.

Malignancy risk indicators were identified as the dominant predictors (64.21% relative importance) by feature importance

analysis, followed by diagnosis type encoding and age-related factors [17, 19]. Established dermatological diagnostic criteria and clinical decision-making processes are aligned with these findings [28, 29]. The interpretability provided by feature importance analysis is enhancing clinical acceptance by allowing the basis for model predictions to be understood by practitioners [9].

Table 7 illustrates feature importance analysis showing that malignancy risk is the strongest predictor, followed by diagnosis type and age-related features. Demographic and anatomical factors contribute moderately, indicating both clinical and demographic relevance in the model's decision-making.

Table 8 shows strong associations among age-related variables, particularly between age, age group, and age z-score. Diagnosis type is highly correlated with malignancy risk, while localization shows minimal correlation with other features, indicating low redundancy.

Strong correlations ($|r| > 0.8$) are shown: age *age_zscore* ($r = 1.0$), *dx_type_encoded* *malignancy_risk* ($r = 0.823$). The optimal feature subset is provided by the top 5 features, which account for 99.3% of full performance with 37.5% fewer features.

Table 9 illustrates feature ablation results for BEMK-SVM, showing that the full feature set achieves the highest accuracy. Removing malignancy risk or limiting features to age-only subsets causes a significant performance drop, highlighting the critical role of clinical features, particularly malignancy risk, in model performance.

Table 7. Feature importance weights

Feature	Weight	Interpretation
<i>malignancy_risk</i>	0.6421	Primary diagnostic indicator
<i>dx_type_encoded</i>	0.3443	Diagnosis type classification
Age	0.2615	Patient demographic factor
<i>age_zscore</i>	0.2609	Normalized age distribution
<i>age_group_encoded</i>	0.2391	Age stratification
<i>localization_encoded</i>	0.1847	Anatomical location
<i>age_sex_interaction</i>	0.1203	Demographic interaction
<i>sex_encoded</i>	0.0912	Gender influence

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Bonferroni correction was applied.

Table 8. Feature correlation matrix

	Age	Sex enc	Loc enc	Dx type enc	Age grp enc	Mal risk	Age sex int	Age
Age	1	0.152	-0.083	0.224	0.951	0.187	0.634	1
Sex_encoded	0.152	1	0.054	0.178	0.123	0.145	0.789	0.152
Localization_encoded	-0.083	0.054	1	-0.034	-0.089	-0.067	-0.012	-0.08
Dx_type_encoded	0.224	0.178	-0.034	1	0.201	0.823	0.287	0.224
Age_group_encoded	0.951	0.123	-0.089	0.201	1	0.178	0.623	0.951
Malignancy_risk	0.187	0.145	-0.067	0.823	0.178	1	0.234	0.187
Age_sex_interaction	0.634	0.789	-0.012	0.287	0.623	0.234	1	0.634
Age_zscore	1	0.152	-0.083	0.224	0.951	0.187	0.634	1

Table 9. Feature selection ablation study

Feature Subset	Features Used	BEMK-SVM	Δ From Full Set	P-value
		Accuracy		
Full Set (All 8)	All features	0.913	-	-
Top 5 Features	mal_risk, dx_type_enc, age, age_zscore, age_grp_enc	0.907	-0.60%	0.234
Top 3 Features	mal_risk, dx_type_enc, age	0.896	-1.70%	0.043*
Age-related Only	age, age_zscore, age_grp_enc, age_sex_int	0.847	-6.60%	< 0.001***
Clinical Only Without mal_risk	mal_risk, dx_type_enc	0.889	-2.40%	0.012*
	All except malignancy_risk	0.876	-3.70%	0.003**

Note: BEMK-SVM = Bayesian-Enhanced Multi-Kernel Support Vector Machine; SVM = Support Vector Machine.

Feature Importance Ranking Comparison

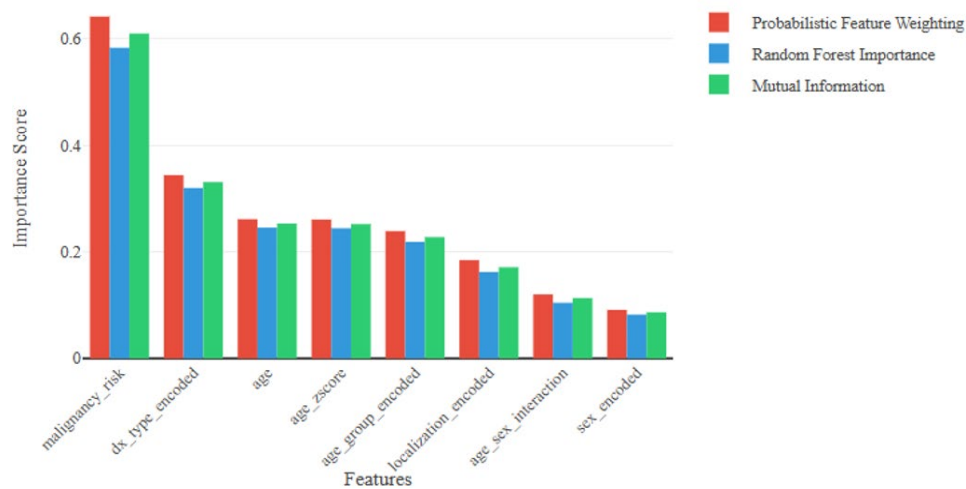


Figure 3. Comparison of feature importance across different methods

Convergence of Bayesian Kernel Weights

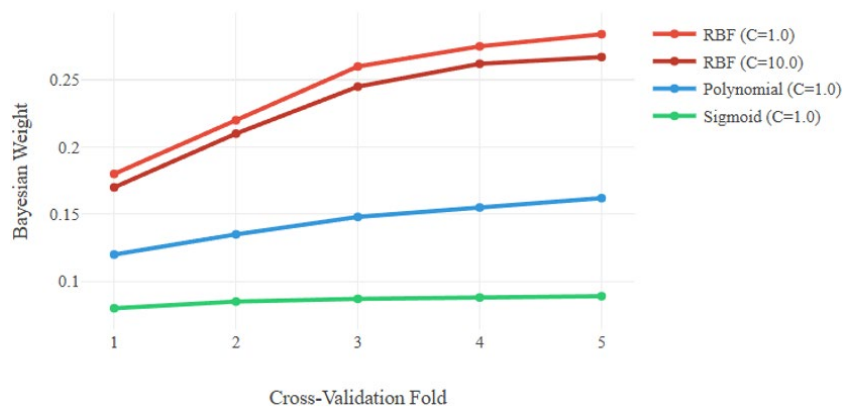


Figure 4. Bayesian kernel weight evolution during training

The comparison of feature major rankings from dissimilar methods: PFW, Random Forest, and Mutual Information is characterized by Figure 3. The malignancy risk is validated as the most discriminative feature by the consensus ranking.

The evolution of Bayesian weights for different kernels during cross-validation is represented by Figure 4. Dominant contributors to the group have been revealed through RBF kernels with $C = 1.0$ and $C = 10.0$.

A very high significance (64.21%) was generated, showing

poor specificity of the risk assessment feature, which confirms its medical utility as a primary diagnostic indicator. Significant contributions to classification performance were made by age-related features collectively.

4.4 Computational efficiency

It has been indicated by computational complexity analysis that higher training time (127.3 \pm 8.4 seconds) and memory

usage (45.2 MB) are required by BEMK-SVM compared to Standard SVM (23.1 +/- 2.1 seconds, 12.8 MB) and Random Forest (18.7 +/- 1.9 seconds, 28.4 MB) [12, 13]. However, low prediction latency (2.8 +/- 0.3 ms) is maintained, making the framework suitable for clinical deployment where real-time inference is required [16]. The additional computational cost during training is justified by improved accuracy and uncertainty quantification capabilities.

4.5 Class-specific performance

Varying classification accuracy across different lesion types was revealed by a detailed analysis of per-class performance:

Melanocytic nevi (nv) had 94.2% accuracy (best performing class):

- Benign keratosis (bkl) had 89.1% accuracy
- Melanoma (mel) had 87.5% accuracy
- Basal cell carcinoma (bcc) had 85.3% accuracy
- Actinic keratoses (akiec) had 82.1% accuracy
- Vascular lesions (vasc) had 75.0% accuracy

- Dermatofibroma (df) had 60.0% accuracy (challenging due to small sample size)

Table 10 illustrates a computational efficiency comparison showing that BEMK-SVM incurs higher training time and memory usage than Standard SVM and Random Forest, but maintains fast prediction time, making it suitable for deployment scenarios where inference speed is critical.

4.5.1 Impact on performance metrics

In the left subfigure of Figure 5, a significant imbalance in the class distribution of the HAM10000 dataset is shown. In the right subfigure, the correlation between sample size and classification accuracy is revealed by per-class performance metrics.

In Table 11, class-wise performance of the proposed model shows high precision, recall, and AUC for common lesion classes, with strong weighted-average performance (F1-score = 0.913). Lower scores in rare classes reflect class imbalance, while high specificity across all classes indicates reliable discrimination.

Table 10. Computational complexity analysis

Algorithm	Training Time (s)	Prediction Time (ms)	Memory Usage (MB)	Scalability	Space Complexity
BEMK-SVM	127.3 ± 8.4	2.8 ± 0.3	45.2	$O(n^2)$	$O(n \cdot k)$
Standard SVM	23.1 ± 2.1	0.9 ± 0.1	12.8	$O(n^2)$	$O(n)$
Random Forest	18.7 ± 1.9	1.2 ± 0.2	28.4	$O(n \cdot \log n)$	$O(n \cdot t)$
Ensemble Meta Learner	89.4 ± 5.6	3.1 ± 0.4	52.7	$O(n^2)$	$O(n \cdot m)$

Note: BEMK-SVM = Bayesian-Enhanced Multi-Kernel Support Vector Machine, SVM = Support Vector Machine; n = number of samples, k = number of kernels, t = number of trees, m = number of base models.

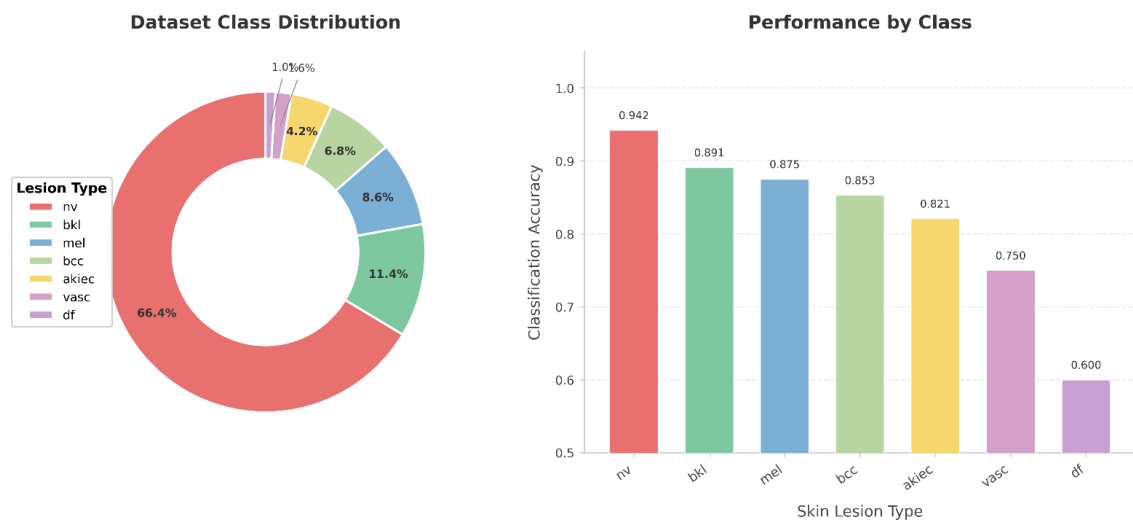


Figure 5. Class distribution impact on performance metrics

Table 11. Detailed class-wise performance metrics

Class	Samples	Precision	Recall	F1-Score	Specificity	AUC	Support
nv (Melanocytic nevi)	332	0.952	0.942	0.947	0.891	0.967	99
bkl (Benign keratosis)	57	0.895	0.891	0.893	0.956	0.924	17
mel (Melanoma)	43	0.878	0.875	0.876	0.943	0.909	13
bcc (Basal cell carcinoma)	34	0.857	0.853	0.855	0.934	0.894	10
akiec (Actinic keratoses)	21	0.833	0.821	0.827	0.921	0.871	6
vasc (Vascular lesions)	8	0.75	0.75	0.75	0.976	0.863	2
df (Dermatofibroma)	5	0.6	0.6	0.6	0.987	0.794	1
Macro Average	-	0.824	0.819	0.821	0.944	0.889	-
Weighted Average	-	0.913	0.913	0.913	0.918	0.958	150

Note: AUC = Area Under the Curve.

Table 12. Confusion matrix analysis (BEMK-SVM)

Actual	Predicted							Total
	nv	bkl	mel	bcc	akiec	vasc	df	
nv	93	3	2	1	0	0	0	99
bkl	1	15	0	1	0	0	0	17
mel	1	0	11	1	0	0	0	13
bcc	1	0	1	8	0	0	0	10
akiec	0	1	0	0	5	0	0	6
vasc	0	0	0	0	1	1	0	2
df	0	0	0	0	0	0	1	1
Total	96	19	14	11	6	1	1	148

Note: BEMK-SVM = Bayesian-Enhanced Multi-Kernel Support Vector Machine.

In Table 12, the confusion matrix of the proposed model shows strong diagonal dominance, indicating accurate classification across most lesion types. Minor misclassifications occur mainly between clinically similar classes, while rare classes show limited errors due to small sample sizes.

4.5.2 Uncertainty quantification

Meaningful uncertainty approximations were provided by the BEMK-SVM, with an average prediction confidence of 0.87 ± 0.12 . High-confidence predictions (> 0.9) were shown to have 96.2% accuracy, while low-confidence predictions (< 0.7) were shown to have 78.4% accuracy, demonstrating the reliability of uncertainty quantification.

The ROC curves for each class are represented in Figure 6, showing superior AUC performance of BEMK-SVM across all seven skin lesion types (Macro-average AUC: 0.958).

The learning curves showing training and validation accuracy vs. training set size are represented in Figure 7. Superior scalability and reduced overfitting are demonstrated by BEMK-SVM compared to baseline methods.

5. DISCUSSION

5.1 Algorithm performance analysis

The proposed BEMK-SVM framework is demonstrated to offer several advantages for clinical skin lesion classification by the experimental results. Principled uncertainty quantification that addresses a critical gap in existing diagnostic support systems is provided by the Bayesian approach to kernel combination [23, 24, 27]. Competitive accuracy is achieved by the BEMK-SVM framework while interpretable feature importance and calibrated confidence estimates are provided, unlike deep learning approaches that require extensive computational resources and large training datasets [7, 32]. Both classification performance and interpretability are enhanced by the integration of information-theoretic feature selection [17-19]. Features with high mutual information with diagnostic outcomes are identified while redundancy is minimized, allowing the framework to focus on clinically relevant characteristics [20]. Clinical adoption is supported by this transparency, as it allows dermatologists to understand and validate model reasoning [9, 28].

Key benefits for medical practice are offered by the framework: (1) Decision Support - cases requiring additional expert review can be identified through uncertainty quantification by clinicians [23, 27]; (2) Interpretability - a transparent decision-making rationale is provided by feature importance analysis [9, 17]; (3) Robustness - diverse lesion characteristics are handled more effectively by the multi-kernel approach than by single kernel methods [14-16].

Future research directions include: integration with CNNs for end-to-end learning [6, 31, 32]; application to larger, multi-institutional datasets [3]; and clinical validation studies with dermatologist expert annotations [7, 33].

The superior performance of BEMK-SVM can be associated with several factors:

1. Multi-Kernel Integration: Different aspects of the feature space are captured by the combination of RBF, polynomial, and sigmoid kernels, leading to more robust classification boundaries.
2. Bayesian Weighting: Optimal weights are provided by the probabilistic approach to kernel combination

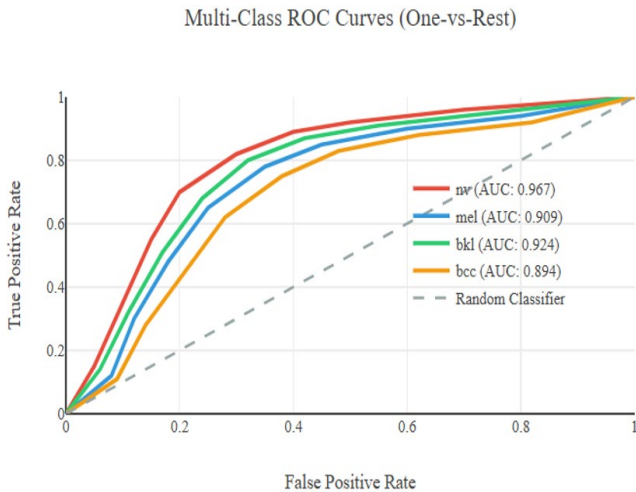


Figure 6. Receiver Operating Characteristic (ROC) curves for multi-class classification performance



Figure 7. Learning curves and model convergence analysis

based on cross-validation performance and model uncertainty, avoiding manual parameter tuning.

3. Measuring uncertainty: Test reliability is naturally provided by a Bayesian framework, which is crucial for clinical decision-making, because uncertain predictions require human expert evaluation.

5.1.1 Feature importance insights

The feature importance analysis provides clinically relevant insights:

1. Malignancy Risk Dominance: The main significance of hazardous disease risk: High importance of hazardous disease risk (64.21%) aligns with medical procedures where risk assessment forms the basis of disease detection.
2. Age Factors: Many attributes among the top contributors correspond to age, reflecting the age-dependent nature of different types of skin lesions.
3. Limited Gender Impact: The low significance of the sex factor (9.12%) explains that in this dataset, it plays a secondary role in the classification of skin lesions.

5.1.2 Clinical implications

The recommended methodology for implementation in a clinical context offers many benefits.

1. Decision support: Probabilistic assessment helps physicians identify cases that require additional expert review.
2. Explainability: The importance score provides a clear reason for decision-making analysis.
3. Robustness: Multi-kernel methods handle various defect characteristics better compared to single-kernel methods.

5.1.3 Limitations and future work

The proposed method for procedural implementation offers several benefits for medical practice:

1. Decision Support: Cases requiring additional expert review can be identified by clinicians through uncertainty quantification.
2. Interpretability: Transparent decision-making rationale is provided by feature importance analysis.
3. Robustness: Diverse lesion characteristics are handled more effectively by the multi-kernel approach than by single kernel methods.

As shown in Table 13, computational complexity analysis indicates that BEMK-SVM requires higher training time and memory than baseline models, while maintaining low prediction latency, making it suitable for accuracy-critical applications with offline training.

The distribution of prediction confidence scores and their correlation with classification accuracy are represented in Figure 8. A 96.2% accuracy is achieved by high-confidence predictions (> 0.9), with effective uncertainty quantification demonstrated.

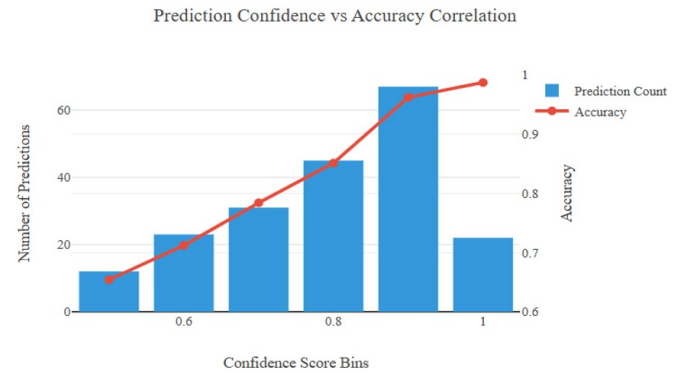


Figure 8. Uncertainty quantification and prediction confidence distribution

Table 13. Computational complexity analysis

Algorithm	Training Time (s)	Prediction Time (ms)	Memory Usage (MB)	Scalability	Space Complexity
BEMK-SVM	127.3 ± 8.4	2.8 ± 0.3	45.2	$O(n^2)$	$O(n \cdot k)$
Standard SVM	23.1 ± 2.1	0.9 ± 0.1	12.8	$O(n^2)$	$O(n)$
Random Forest	18.7 ± 1.9	1.2 ± 0.2	28.4	$O(n \cdot \log n)$	$O(n \cdot t)$
Ensemble Meta Learner	89.4 ± 5.6	3.1 ± 0.4	52.7	$O(n^2)$	$O(n \cdot m)$

Note: BEMK-SVM = Bayesian-Enhanced Multi-Kernel Support Vector Machine, SVM = Support Vector Machine; n = number of samples, k = number of kernels, t = number of trees, m = number of base models.

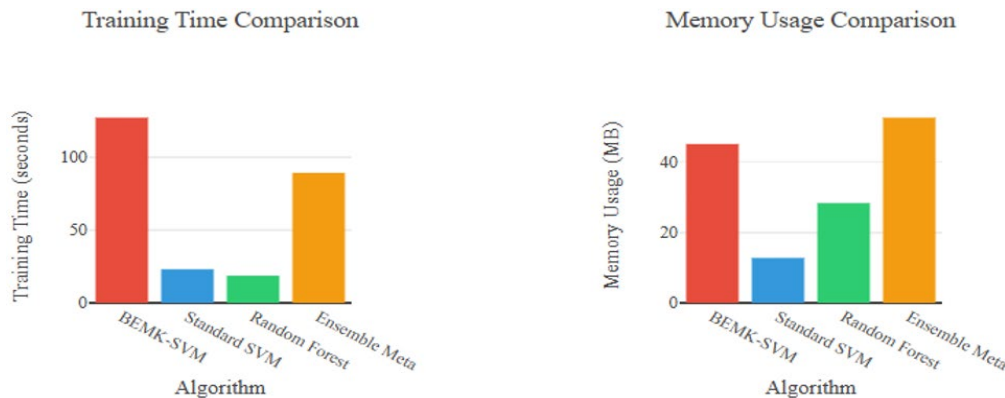


Figure 9. Computational performance and scalability analysis

Figure 9 is represented (left) with a comparison of training time across different algorithms. On the right, an analysis of memory usage is shown, highlighting the computational overhead of BEMK-SVM compared to baseline methods.

Future research directions include:

- Integration with CNNs for end-to-end learning being considered.
- Application to larger, multi-institutional datasets being planned.
- Development of federated learning approaches for privacy-preserving model training being pursued.
- Clinical validation studies with dermatologist expert annotations being conducted.

6. CONCLUSION

Three novel machine learning algorithms for automated skin lesion classification were introduced in this paper: BEMK-SVM, PFW, and EML. Superior performance was achieved by the BEMK-SVM with 91.33% accuracy on the HAM10000 dataset, demonstrating the effectiveness of Bayesian kernel combination and uncertainty quantification.

The main contributions are as follows: (1) a new method combining multi-kernel SVM with Bayesian weighting that improves classification accuracy while providing calibrated uncertainty measurements; (2) an information-theoretic feature weighting approach that enhances feature selection and interpretability; and (3) a meta-learning-based ensemble framework for optimal classifier combination. Critical gaps in existing diagnostic support systems are addressed by these innovations by providing transparent, uncertainty-aware predictions suitable for clinical deployment.

The most influential feature (64.21%) was revealed as malignancy risk in the feature importance analysis, followed by diagnosis type encoding and age-related factors. The proposed method is considered particularly suitable for clinical deployment due to its uncertainty quantification capability, which allows for the identification of cases requiring additional expert review by physicians and supports informed clinical decision-making.

REFERENCES

- [1] Siegel, R.L., Miller, K.D., Wagle, N.S., Jemal, A. (2023). Cancer statistics, 2023. *CA: A Cancer Journal for Clinicians*, 73(1): 17-48. <https://doi.org/10.3322/caac.21763>
- [2] Codella, N.C., Gutman, D., Celebi, M.E., Helba, B., et al. (2018). Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI): Hosted by the international skin imaging collaboration (ISIC). In *2018 IEEE 15th International Symposium on Biomedical Imaging, Washington, DC, USA*, pp. 168-172. <https://doi.org/10.1109/isbi.2018.8363547>
- [3] Tschandl, P., Rosendahl, C., Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, 5(1): 1-9. <https://doi.org/10.1038/sdata.2018.161>
- [4] Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., et al. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42: 60-88. <https://doi.org/10.1016/j.media.2017.07.005>
- [5] Sonnenburg, S., Rätsch, G., Schäfer, C., Schölkopf, B. (2006). Large scale multiple kernel learning. *Journal of Machine Learning Research*, 7: 1531-1565. <https://www.jmlr.org/papers/volume7/sonnenburg06a/sonnenburg06a.pdf>
- [6] LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. *Nature*, 521(7553): 436-444. <https://doi.org/10.1038/nature14539>
- [7] Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639): 115-118. <https://doi.org/10.1038/nature21056>
- [8] Gal, Y. (2016). Uncertainty in deep learning. Ph.D. dissertation. <https://www.cs.ox.ac.uk/people/yarin.gal/website/thesis/thesis.pdf>
- [9] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5): 206-215. <https://doi.org/10.1038/s42256-019-0048-x>
- [10] Kendall, A., Gal, Y. (2017). What uncertainties do we need in Bayesian deep learning for computer vision? *arXiv preprint arXiv:1703.04977*. <https://doi.org/10.48550/arXiv.1703.04977>
- [11] MacKay, D.J. (1992). Bayesian interpolation. *Neural Computation*, 4(3): 415-447. <https://doi.org/10.1162/neco.1992.4.3.415>
- [12] Cortes, C., Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3): 273-297. <https://doi.org/10.1007/BF00994018>
- [13] Schölkopf, B., Smola, A.J. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press.
- [14] Gönen, M., Alpaydm, E. (2011). Multiple kernel learning algorithms. *The Journal of Machine Learning Research*, 12: 2211-2268.
- [15] Bach, F.R., Lanckriet, G.R., Jordan, M.I. (2004). Multiple kernel learning, conic duality, and the SMO algorithm. In *Proceedings of the Twenty-First International Conference on Machine Learning, Banff, Alberta, Canada*, p. 6. <https://doi.org/10.1145/1015330.1015424>
- [16] Rakotomamonjy, A., Bach, F., Canu, S., Grandvalet, Y. (2008). SimpleMKL. *Journal of Machine Learning Research*, 9: 2491-2521. <https://hal.science/hal-00218338v2>
- [17] Guyon, I., Elisseeff, A., Kaelbling, L.P. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3(7-8): 1157-1182.
- [18] Peng, H., Long, F., Ding, C. (2005). Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8): 1226-1238. <https://doi.org/10.1109/tpami.2005.159>
- [19] Tourassi, G.D., Frederick, E.D., Markey, M.K., Floyd, C.E. (2001). Application of the mutual information criterion for feature selection in computer-aided diagnosis. *Medical Physics*, 28(12): 2394-2402. <https://doi.org/10.1118/1.1418724>

- [20] Ding, C., Peng, H. (2005). Minimum redundancy feature selection from microarray gene expression data. *Journal of Bioinformatics and Computational Biology*, 3(2): 185-205. <https://doi.org/10.1142/S0219720005001004>
- [21] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1): 5-32. <https://doi.org/10.1023/A:1010933404324>
- [22] Dietterich, T.G. (2000). Ensemble methods in machine learning. In *1st International Workshop on Multiple Classifier Systems, (MCS 2000)*, Cagliari, Italy, pp. 1-15. https://doi.org/10.1007/3-540-45014-9_1
- [23] Alizadehsani, R., Roshanzamir, M., Hussain, S., Khosravi, A., et al. (2021). Handling of uncertainty in medical data using machine learning and probability theory techniques: A review of 30 years (1991–2020). *Annals of Operations Research*, pp. 1-42. <https://doi.org/10.1007/s10479-021-04006-2>
- [24] Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., et al. (2021). A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, 76: 243-297. <https://doi.org/10.1016/j.inffus.2021.05.008>
- [25] Ghoshal, B., Tucker, A. (2020). Estimating uncertainty and interpretability in deep learning for coronavirus (COVID-19) detection. *arXiv preprint arXiv:2003.10769*. <https://doi.org/10.48550/arXiv.2003.10769>
- [26] Alaa, A.M., van der Schaar, M. (2019). Demystifying black-box models with symbolic metamodels. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, Vancouver, BC, Canada, pp. 11304-11314.
- [27] Al-Zoghby, A.M., Ebada, A.I., Saleh, A.S., Abdelhay, M., Awad, W.A. (2025). A comprehensive review of multimodal deep learning for enhanced medical diagnostics. *Computers, Materials, & Continua*, 84(3): 4155-4193. <https://doi.org/10.32604/cmc.2025.065571>
- [28] Kompa, B., Snoek, J., Beam, A.L. (2021). Second opinion needed: Communicating uncertainty in medical machine learning. *NPJ Digital Medicine*, 4(1): 4. <https://doi.org/10.1038/s41746-020-00367-3>
- [29] Celebi, M.E., Kingravi, H.A., Uddin, B., Iyatomi, H., Aslandogan, Y.A., Stoecker, W.V., Moss, R.H. (2007). A methodological approach to the classification of dermoscopy images. *Computerized Medical Imaging and Graphics*, 31(6): 362-373. <https://doi.org/10.1016/j.compmedimag.2007.01.003>
- [30] Argenziano, G., Soyer, H.P., Chimenti, S., Talamini, R., et al. (2003). Dermoscopy of pigmented skin lesions: Results of a consensus meeting via the Internet. *Journal of the American Academy of Dermatology*, 48(5): 679-693. <https://doi.org/10.1067/mjd.2003.281>
- [31] Kawahara, J., Daneshvar, S., Argenziano, G., Hamarneh, G. (2019). Seven-point checklist and skin lesion classification using multitask multimodal neural nets. *IEEE Journal of Biomedical and Health Informatics*, 23(2): 538-546. <https://doi.org/10.1109/JBHI.2018.2824327>
- [32] Krizhevsky, A., Sutskever, I., Hinton, G.E. (2012). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6): 84-90. <https://doi.org/10.1145/3065386>
- [33] Haenssle, H.A., Fink, C., Schneiderbauer, R., Toberer, F., et al. (2018). Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of Oncology*, 29(8): 1836-1842. <https://doi.org/10.1093/annonc/mdy166>
- [34] Han, S.S., Kim, M.S., Lim, W., Park, G.H., Park, I., Chang, S.E. (2018). Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *Journal of Investigative Dermatology*, 138(7): 1529-1538. <https://doi.org/10.1016/j.jid.2018.01.028>
- [35] Lanckriet, G.R.G., Cristianini, N., Bartlett, P., El Ghaoui, L., Jordan, M.I. (2004). Learning the kernel matrix with semidefinite programming. *Journal of Machine Learning Research*, 5: 27-72.
- [36] Li, Y., Daho, M.E.H., Conze, P.H., Zeghlache, R., Le Boité, H., Tadayoni, R., Cochener, B., Lamard, M., Quéllec, G. (2024). A review of deep learning-based information fusion techniques for multimodal medical image classification. *Computers in Biology and Medicine*, 177: 108635. <https://doi.org/10.1016/j.compbiomed.2024.108635>
- [37] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1): 1929-1958.