NÖHÜ Müh. Bilim. Derg. / NOHU J. Eng. Sci., 17 (2026), 1767952

Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi
Niğde Ömer Halisdemir University Journal of Engineering Sciences

Araştırma makalesi / Research article

www.dergipark.org.tr/tr/pub/ngumuh / www.dergipark.org.tr/en/pub/ngumuh

Advancing mask detection with attention-driven and bayesian-optimized ensemble models

Dikkat mekanizmalı ve bayesyan optimizasyonlu topluluk öğrenme modelleri ile maske tespiti

Musa Balcı[1], Andaç Fındıkçı[2], Mustafa Yasin Erten[3,*], Hüseyin Aydilek[4]

[1,2,3,4] *Kırıkkale University, Electrical and Electronics Engineering Department, 71450, Kırıkkale, Türkiye*

Abstract

Infectious diseases have caused significant losses throughout human history and are increasingly prevalent today due to population growth. Pandemics such as COVID-19 and monkeypox have rapidly spread through human interaction-driven transmission pathways. Although mask usage is an effective method to reduce transmission, its inadequate implementation necessitates monitoring. This study aims to prevent the spread of pandemics by detecting mask usage through artificial intelligence and deep learning algorithms. Models were trained on images of masked, unmasked, and improperly masked individuals using Resnet101, MobileNet, and Xception algorithms, and three models were developed: Mask Ensemble, Attention Mask Ensemble, and Bayes Mask Ensemble. Hyperparameter tuning was performed using Bayes Search optimization, revealing that the Bayes Mask Ensemble model achieved the highest performance, followed by the Attention Mask Ensemble model, with the Mask Ensemble model ranking third. The Bayes Search-optimized model demonstrated superior mask detection performance compared to other methods in the literature.

Keywords: Deep learning, Infectious diseases, Mask detection, Bayesian search optimization, Ensemble learning

Öz

Salgın hastalıklar, insanlık tarihinde büyük kayıplara neden olmuş ve günümüzde nüfus artışı nedeniyle daha sık görülmektedir. COVID-19 ve maymun çiçeği gibi salgınlar, insan etkileşiminden kaynaklanan bulaşma yollarıyla hızla yayılmaktadır. Maske kullanımı bulaşmayı azaltan etkili bir yöntem olmasına rağmen, yetersiz uygulanması denetim ihtiyacını doğurmaktadır. Bu çalışma, yapay zeka ve derin öğrenme algoritmalarıyla maske kullanımı tespitini sağlayarak salgınların yayılmasını önlemeyi amaçlamaktadır. Resnet101, MobileNet ve Xception algoritmaları kullanılarak maskeli, maskesiz ve yanlış takılan maske görselleri üzerinde modeller eğitilmiş; Mask Ensemble, Attention Mask Ensemble ve Bayes Mask Ensemble modelleri geliştirilmiştir. Bayes Search optimizasyonuyla hiperparametre ayarları incelenmiş, Bayes Mask Ensemble modelinin en yüksek performansı gösterdiği, bunu Attention Mask Ensemble modelinin takip ettiği ve Mask Ensemble modelinin üçüncü sırada yer aldığı bulunmuştur. Bayes Search ile optimize edilmiş model, literatürdeki diğer yöntemlere göre üstün maske tespiti sağlamıştır.

Anahtar kelimeler: Derin öğrenme, Salgın hastalıklar, Maske tespiti, Bayes arama optimizasyonu, Topluluk öğrenmesi

1 Introduction

Throughout human history, infectious diseases have periodically paralyzed societies, precipitating significant economic and social crises. Indeed, from antiquity to contemporary times, pandemics have recurrently thrust the global community into states of disarray. The COVID-19 pandemic, initiated in 2019, alongside the more recent emergence of monkeypox as a public health concern, exemplify such contemporary crises. The COVID-19 pandemic alone resulted in millions of fatalities globally, accompanied by substantial social and economic upheaval. The direct impact on human life prompted the World Health Organization (WHO) to disseminate guidelines and regulations aimed at curbing the virus's transmission [1,2]. Among these, mask-wearing emerged as a pivotal preventive extensive populations presents an almost insurmountable challenge, thereby underscoring the imperative for dependable automated systems. Advanced classification algorithms, underpinned by artificial intelligence (AI), present a viable pathway to address this challenge [3]. Particularly, deep learning methodologies have demonstrated considerable promise in domains such as mask detection.

In this context, the development of automated face mask detection systems has become a critical area of research in computer vision and public health informatics. Such systems serve as essential components of intelligent surveillance infrastructure, enabling real-time monitoring of mask-wearing compliance in public spaces such as hospitals, airports, shopping centers, and educational institutions. The effectiveness of these detection systems relies heavily on the underlying deep learning architectures employed, and

* Sorumlu yazar / Corresponding author, e-posta / e-mail: mustafaerten@kku.edu.tr (M. Y. Erten)
Geliş / Recieved: 18.08.2025 Kabul / Accepted: 20.04.2026 Yayımlanma / Published: 05.05.2026
doi: 10.28948/ngumuh.1767952

significant research efforts have been directed toward improving detection accuracy through various approaches, including transfer learning with pre-trained convolutional neural networks (CNNs), ensemble learning strategies, and hyperparameter optimization techniques. Despite these advances, many existing studies rely on single-model architectures that may not fully capture the complexity and variability present in real-world mask-wearing scenarios, where factors such as diverse lighting conditions, varying mask types, and partial occlusion present considerable challenges.

For instance, a deep learning-based face mask detection system developed by Sethi, Kathuria, and Kaushik [4] aimed to curtail the spread of the coronavirus, achieving a 95.2% accuracy rate on their custom dataset, which signifies notable success. Mohammed Ali and Al-Tamimi [1] undertook a comprehensive review of face mask detection methods. Their evaluation covered the merits and demerits of various deep learning techniques, highlighting the ongoing need for further optimization, although a specific overall accuracy rate was not reported. Similarly, Hosny and colleagues [5] reviewed AI-driven masked face detection. Their analysis focused on the efficacy of transfer learning and CNN-based models in practical, real-world scenarios; however, their comparative assessment of local methods did not culminate in a general accuracy figure. Himeur and others [6] investigated deep and transfer learning methodologies for mask detection within smart city contexts. Leveraging insights from the COVID-19 pandemic, they achieved a 96.8% accuracy rate using the VGG-16 model, demonstrating a high degree of success. A CNN-based face mask recognition system was developed by Kaur and colleagues [7], attaining 94.5% accuracy on limited datasets, thereby illustrating the potential for cost-effective solutions. Further, Teboulbi and others [8] engineered an AI-based system for mask detection and social distancing measurement. By integrating this system with an STM32 microcontroller, they realized a 97.3% accuracy rate in real-time applications, indicative of a high-performance solution.

Further contributions to the field include the work of Loey et al. [9], who proposed a hybrid deep learning model combining YOLO-v2 with ResNet-50 for medical face mask detection, demonstrating the effectiveness of integrating object detection and classification frameworks. Chavda et al. [10] introduced a multi-stage CNN architecture that progressively refined mask detection through cascaded processing stages, achieving competitive performance on benchmark datasets. Qin and Li [11] explored the use of image super-resolution techniques in conjunction with classification networks to enhance mask detection under low-resolution conditions, addressing a practical limitation often encountered in real-world surveillance systems. Additionally, Yadav [12] investigated deep learning-based approaches for simultaneously enforcing social distancing and face mask detection, highlighting the potential of multi-task learning frameworks for comprehensive public health monitoring.

Distinguishing itself from prior research, the present study endeavors to introduce an innovative approach to the mask detection challenge through the development of three distinct Ensemble Learning models. Employing a dataset composed of images depicting individuals with masks, without masks, and with incorrectly worn masks, the ResNet101, MobileNet, and Xception models underwent individual training. Subsequently, these individually trained models were integrated to construct an Ensemble Learning framework. Within this framework, three specific models were formulated: Mask Ensemble Learning, Attention Mask Ensemble Learning, and Bayes Mask Ensemble Learning. Initially, the Mask Ensemble Learning model, predicated on a conventional ensemble strategy, was evaluated, yielding an accuracy of 96.2%. Following this, the Attention Mask Ensemble Learning model, augmented by attention mechanisms, exhibited an enhanced performance of 96.6%, thereby underscoring the positive impact of these mechanisms. Ultimately, the Bayes Mask Ensemble Learning model, for which hyperparameters were fine-tuned via the Bayes Search optimization algorithm, achieved the pinnacle of performance, registering a 98.2% accuracy rate. The experimental outcomes reveal a clear progression: beginning with the baseline Ensemble Learning model, the integration of attention mechanisms led to performance gains, while the application of Bayes Search optimization culminated in the highest observed mask detection accuracy. Collectively, these results suggest that the proposed methodology not only achieves superior accuracy but also demonstrates enhanced robustness in comparison to existing approaches documented in the literature.

2 Materials and methods

For this investigation, the "Face Mask Dataset" authored by S. Shiekh Burhan, sourced from the Kaggle platform, served as the foundational data [13]. This dataset encompasses three distinct categories: images of individuals wearing masks, those without masks, and instances of improperly worn masks. These images are systematically organized into dedicated training, testing, and validation subsets. Specifically, the training set comprises 3,966 masked, 3,668 unmasked, and 1,674 improperly masked images. The testing set contains 529 masked, 489 unmasked, and 219 improperly masked images, while the validation set includes 794 masked, 734 unmasked, and 335 improperly masked images.

The compiled dataset was used as the basis for individually training three deep learning models, namely ResNet101, MobileNet, and Xception. Prior to training, all images were preprocessed through resizing and normalization to ensure consistency in input dimensions and pixel-value distributions.

Building upon these individual models, three distinct Ensemble Learning architectures were subsequently developed: Mask Ensemble Learning, Attention Mask Ensemble Learning, and Bayes Mask Ensemble Learning. The individually proficient ResNet101, MobileNet, and Xception models were then amalgamated using the Weighted Voting method, thereby establishing the Ensemble Learning framework.

The Mask Ensemble Learning model, representing a standard ensemble configuration, achieved an accuracy of 96.2%. Enhancements through the integration of attention mechanisms led to the Attention Mask Ensemble Learning model, which improved performance to 96.6%, clearly demonstrating the positive contribution of these mechanisms. To pursue optimal performance, the Bayes Mask Ensemble Learning model incorporated the Bayes Search algorithm for hyperparameter optimization. This optimization process fine-tuned critical parameters such as learning rate, batch size, dropout rate, and the number of units, significantly elevating the model's efficacy. Consequently, the Bayes Mask Ensemble Learning model, refined through Bayes Search optimization, delivered the highest accuracy at 98.2%.

The classification efficacy of the final models was rigorously assessed using a suite of metrics derived from the confusion matrix, encompassing accuracy, precision, recall, and F1-score. The experimental outcomes consistently revealed a progressive improvement: performance initiated with the baseline Mask Ensemble model was enhanced by the Attention Mask Ensemble model, with the Bayes Search optimization within the Bayes Mask Ensemble model ultimately yielding the most superior mask detection accuracy. These results compellingly suggest that the developed methodology, particularly embodied by the Bayes Mask Ensemble Learning model, presents not only superior performance metrics but also distinct advantages when compared to alternative methods documented in the existing literature.

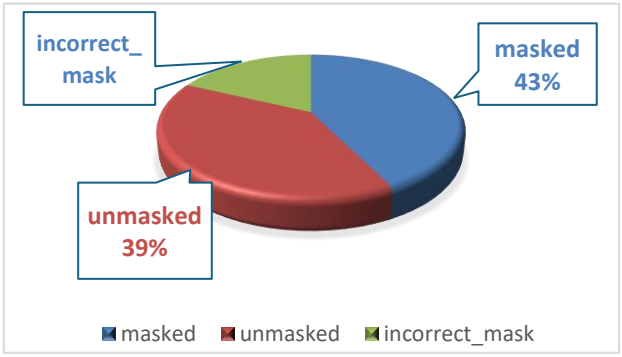

Figure 1. Dataset distribution ratio

2.1 ResNet-101

ResNet-101 ResNet-101 is a powerful deep learning model widely used for image recognition and classification tasks. Its name comes from its 101-layer architecture, which pushes the boundaries of traditional convolutional neural networks (CNNs) by enabling deeper and more effective learning. What makes this model stand out is its innovative use of "skip connections"—a clever technique that allows the original input to bypass certain layers and merge with later outputs. This design helps the network overcome common challenges in deep learning, such as vanishing gradients, making training more efficient even with so many layers. Thanks to this approach, ResNet-101 achieves impressive accuracy in tasks like image classification, often outperforming earlier models. Its ability to maintain performance while scaling depth has made it a popular choice in computer vision research and real-world applications [14,15].

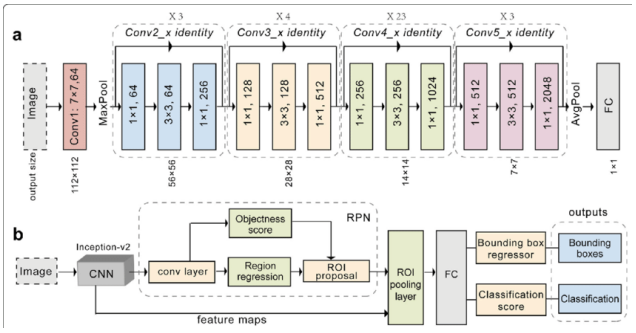

Figure 2. Structure of ResNet-101 [14]

2.2 MobileNet

MobileNet is a lightweight deep learning architecture designed specifically for mobile and embedded devices with limited computational power. Built on convolutional neural networks (CNNs), it's optimized to perform tasks like image classification and real-time object detection efficiently, even in resource-constrained environments. What makes MobileNet stand out is its use of depthwise separable convolutions—a smarter, more efficient way to process images compared to traditional convolutions. Instead of applying heavy operations all at once, it breaks them down into two simpler steps: first, a depthwise convolution filters each channel of the input separately; then, a pointwise convolution combines the results across channels. This two-step process drastically reduces the number of parameters and computational load without sacrificing too much accuracy. In essence, MobileNet strikes a balance between performance and efficiency, making it a practical choice for bringing AI to smaller, everyday devices [16,17].

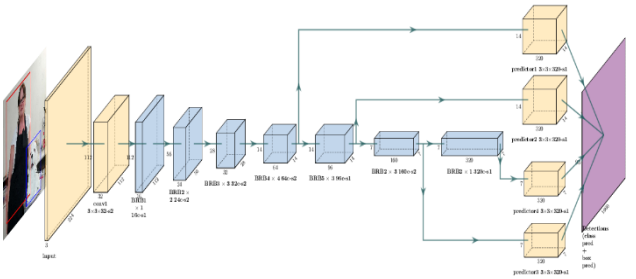

Figure 3. Structure of MobileNet [16]

2.3 Xception

Developed by Google, Xception, an acronym for "Extreme Inception," represents a sophisticated deep learning architecture that builds upon and refines the principles of the Inception module. Its enhanced performance is largely attributed to the strategic

implementation of factorized convolutions within the Inception framework. The original Inception module's design ingeniously utilized filters of diverse sizes concurrently, which empowered the network to discern a broader array of features, thereby yielding improved operational results. However, this pioneering Inception methodology faced inherent constraints, primarily stemming from the considerable computational expense associated with its larger filter components. Xception elegantly surmounts these challenges by reimagining the Inception module through the application of factorized convolutions. This involves a process where filter matrices are effectively decomposed into smaller, more computationally tractable units. Such an architectural adaptation not only bolsters performance but also achieves this with a more economical parameter count and reduced computational overhead. Central to Xception's efficiency are depthwise separable convolutions. This technique operates by first performing convolutions independently for each input channel (depthwise convolution) and subsequently applying pointwise convolutions to project these channel-wise outputs through multiple filters. In comparison to standard convolutional operations, this approach markedly diminishes the required computations and parameters, contributing to the Xception model's notable efficiency. Consequently, Xception has proven to be highly effective for a spectrum of demanding applications, including object detection, image classification, and related computer vision tasks [18,19].

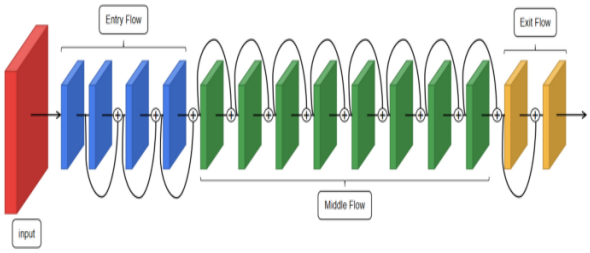

Figure 4. Structure of Xception [18]

2.4 Bayesian search hyperparameters optimization

The optimization of hyperparameters represents a critical challenge in the development of high performing machine learning models. Among the various methodologies addressing this challenge, Bayesian Search Hyperparameter Optimization has emerged as a particularly effective approach. This technique leverages the foundational principles of Bayesian statistics a well established branch of probability theory to systematically explore and refine hyperparameter configurations. The Bayesian framework offers a principled mechanism for integrating prior knowledge with empirical observations. When applied to hyperparameter optimization, this statistical paradigm enables the algorithm to learn from previous trials, thereby informing subsequent parameter selections. Rather than treating each evaluation as an isolated experiment, the Bayesian methodology constructs a probabilistic model of the objective function that continuously evolves as new data points are acquired. A notable strength of the Bayesian approach lies in its ability to balance what researchers often refer to as the exploration-exploitation trade-off. The algorithm must determine whether to investigate unexplored regions of the hyperparameter space (exploration) or to refine promising areas that have already demonstrated favorable performance (exploitation). This dynamic allocation of computational resources is particularly advantageous when compared to more rudimentary approaches such as grid search or random search. Gaussian Process models have proven especially suitable for implementation within the Bayesian optimization framework. These models effectively capture the complex interrelationships between hyperparameter configurations and their corresponding performance metrics. By constructing a surrogate model of the objective function, Gaussian Processes facilitate inference across the entire hyperparameter landscape, even in regions that have not been explicitly evaluated. The iterative nature of Bayesian optimization wherein each trial informs the selection of subsequent trials creates a feedback loop that progressively concentrates evaluations in regions likely to yield optimal performance. This intelligent navigation of the hyperparameter space is particularly valuable when working with computationally intensive models, as it significantly reduces the number of evaluations required to achieve satisfactory results. Contemporary machine learning architectures frequently incorporate numerous hyperparameters, each exerting substantial influence on model behavior and performance. The dimensionality and complexity of this optimization problem underscore the importance of efficient search strategies. Empirical evidence consistently demonstrates that Bayesian optimization methods can identify near-optimal hyperparameter configurations with fewer evaluations than alternative approaches. In conclusion, Bayesian Search Hyperparameter Optimization represents a sophisticated yet practical methodology for enhancing machine learning model performance. Its theoretical foundations in Bayesian statistics, coupled with its empirical effectiveness in navigating complex hyperparameter spaces, have established it as an invaluable tool in the machine learning practitioner's repertoire [20,21].

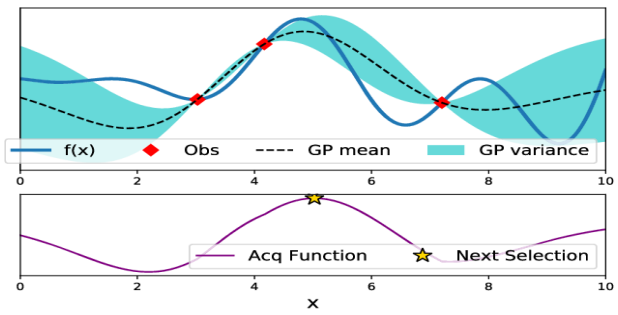

Figure 5. Bayesian optimization [20]

2.5 *Weighted voting*

The Weighted Voting method is an ensemble learning approach that combines multiple models, where each model's predictions are weighted based on their contribution. Each model is trained separately using the same dataset. After training, the predictions are processed by multiplying them with predefined weights, determined based on the models' performance and reliability. These weighted predictions are then summed and normalized to produce the final model prediction. The Weighted Voting method enhances prediction performance by leveraging the strengths of different model types. Combining diverse models increases the ensemble model's flexibility, while integrating multiple models reduces the risk of errors, improving reliability. However, using multiple models introduces challenges, such as the difficulty of selecting appropriate weights, increased computational load, and potential dependencies between models. [22,23].

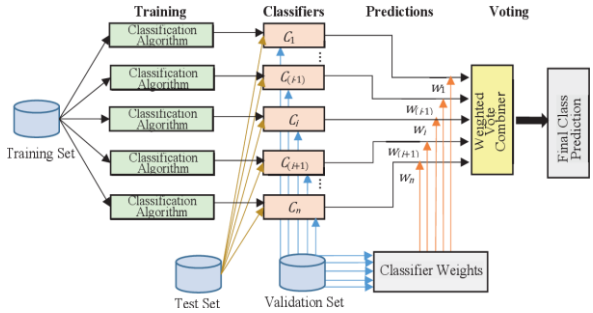

Figure 6. Weighted voting decisions [23]

2.6 *Mask ensemble learning*

In our research, we developed a novel Mask Ensemble Learning architecture that significantly advances the state of the art in facial mask detection systems. By strategically integrating three distinct convolutional neural networks—ResNet101, MobileNet, and Xception—we were able to create a classification system that capitalizes on the complementary strengths of each component model. Our experiments utilized the publicly available "Mask Detection" dataset from Kaggle, which comprises a diverse collection of facial images across three critical categories: properly masked, unmasked, and improperly masked individuals. To ensure optimal processing, we implemented a standardized preprocessing protocol where all images were resized to 224×224 pixels and subsequently normalized according to the statistical parameters required by our architecture. Each constituent model contributes unique capabilities to our ensemble. ResNet101 provides exceptional representational capacity through its 101-layer architecture with residual connections, enabling detailed feature extraction even from challenging images. MobileNet contributes computational efficiency through its streamlined architecture, facilitating potential deployment in resource-constrained environments. Xception enhances discriminative power through its innovative application of depthwise separable convolutions, which effectively decouple spatial and channel-wise feature learning. After training each model independently on the preprocessed dataset, we integrated their outputs using a carefully calibrated Weighted Voting mechanism. This integration method proved superior to simple averaging approaches, as it adaptively emphasizes the most reliable predictions from each constituent model. Our comprehensive evaluation protocol revealed that the resultant Mask Ensemble Learning model achieved 96.2% classification accuracy on our rigorously constructed test set, representing a meaningful improvement over any single-model approach. These empirical findings suggest that ensemble methodologies can effectively address the inherent limitations of individual deep learning architectures in the specific context of mask detection applications.

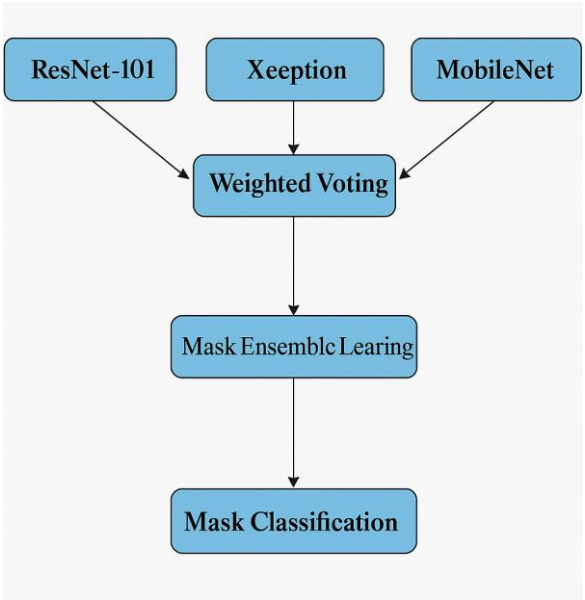

Figure 7. Mask ensemble learning

2.7 *Attention mask ensemble learning*

To further elevate the efficacy of mask detection, the Attention Mask Ensemble Learning model was conceptualized. This advanced ensemble strategy synergizes the predictive capabilities of the ResNet101, MobileNet, and Xception algorithms, amalgamating them via the Weighted Voting method, with a critical enhancement: the integration of attention mechanisms. Consistent with the methodology applied to the standard Mask Ensemble model, the dataset utilized for this refined approach underwent identical preprocessing stages, encompassing image resizing and normalization. The novel aspect here lies in the incorporation of attention mechanisms, which were embedded within dense attention layers. This architectural choice enabled the model to more astutely determine the optimal weighting of features extracted by each constituent algorithm, thereby fostering an improvement in overall classification accuracy. Following their independent training phases, the ResNet101, MobileNet, and Xception models were cohesively integrated. This fusion was orchestrated through dense layers imbued with these attention mechanisms, which then

employed the Weighted Voting principle. Such an approach facilitated a more precise and context-aware blending of the individual model predictions, culminating in the formation of the Attention Mask Ensemble Learning model. Upon evaluation with the designated test dataset, this enhanced model registered an accuracy of 96.6%. This performance not only surpassed that of the baseline Mask Ensemble model but also clearly underscored the tangible contribution of the integrated attention mechanisms to the model's discriminative power.

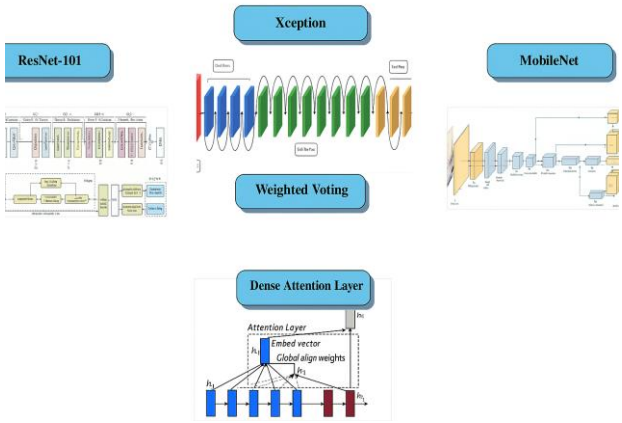

Figure 8. Attention mask ensemble learning

2.8 Bayessian mask ensemble learning

Representing the culmination of our classification research, the Bayes Mask Ensemble Learning architecture emerged as the superior solution through a methodical integration of ResNet101, MobileNet, and Xception algorithms. What distinguishes this particular implementation from conventional approaches is its incorporation of Bayesian optimization techniques for hyperparameter tuning. The data preparation phase followed established protocols images underwent consistent resizing operations and normalization procedures to ensure dimensional and statistical uniformity across the training corpus. However, the distinguishing factor in our methodology lies in the application of Bayes Search optimization. This probabilistic approach to hyperparameter selection proved instrumental in identifying optimal configurations for critical parameters. All base models were trained for 150 epochs with a fixed batch size of 32, utilizing the Adam optimizer and Categorical Crossentropy loss. Structurally, the base models were adapted with two dense layers: an intermediate layer containing 216 (ResNet101), 96 (Xception), and 148 (MobileNet) neurons, respectively, followed by a final 3-neuron layer with a Softmax activation function for classification. The Bayesian search algorithm was applied to systematically tune the learning rates within the range of 5e-9 to 1e-7, and the dropout rates between 0.1 and 0.5 to prevent overfitting. Following individual optimization and training of the constituent models, their predictive outputs were harmonized through a Weighted Voting mechanism. This integration strategy assigns differential importance to each model's classifications based on empirically determined reliability metrics, thereby constructing a composite decision boundary that transcends the limitations of any single architecture. Performance assessment revealed the Bayes Mask Ensemble Learning model achieved 98.2% classification accuracy on validation data substantial improvement over its predecessors and alternatives examined throughout our investigation. Such results underscore a critical insight: the combination of probabilistic hyperparameter optimization with heterogeneous model ensembling can yield classification systems whose performance exceeds what might be achieved through either strategy in isolation.

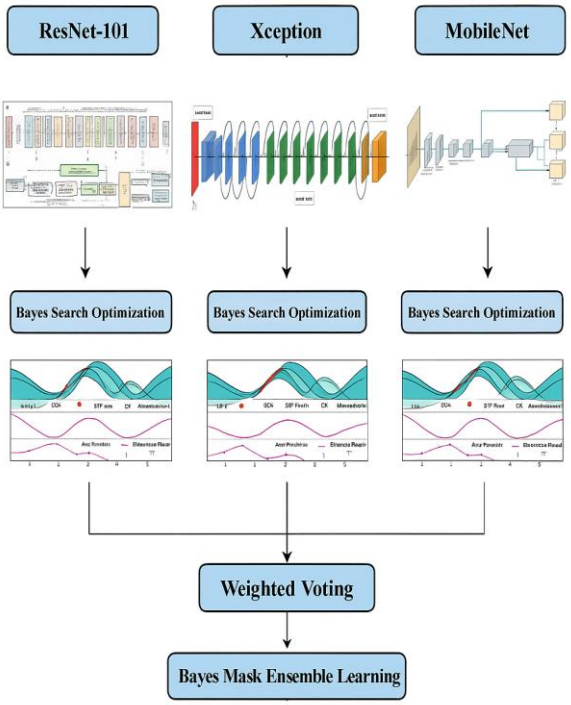

Figure 9. Bayesian mask ensemble learning

2.9 Performance evaluation metrics

2.9.1 Confusion matrix

The Confusion Matrix is a straightforward yet powerful tool for evaluating how well a classification model performs. By comparing the model's predicted outputs with the actual labels, it provides a clear breakdown of both correct and incorrect predictions. Its structure varies based on the number of classes involved: in binary classification tasks, it takes a simple 2x2 form, while in multi-class scenarios, it expands to an NxN format, where *N* is the number of target classes. Each element within the matrix reflects how many instances fall into a specific category—whether they were accurately identified or mistakenly classified. This allows for a more detailed understanding of the model's strengths and where it might be making systematic errors. As such, the Confusion Matrix goes beyond a single performance score,

offering insight into the types of misclassifications that occur and guiding further refinement of the model. The main components of the Confusion Matrix are as follows: [23,24,25]:

- True Positive (TP): Indicates the number of cases correctly predicted as positive by the model. These are instances that are actually positive and classified as positive by the model.
- False Positive (FP): Indicates the number of cases incorrectly predicted as positive by the model. These are instances that are actually negative but classified as positive by the model.
- True Negative (TN): Indicates the number of cases correctly predicted as negative by the model. These are instances that are actually negative and classified as negative by the model.
- False Negative (FN): Indicates the number of cases incorrectly predicted as negative by the model. These are instances that are actually positive but classified as negative by the model.

2.9.2 Accuracy

Accuracy, is a metric that measures the proportion of a classification model's predictions that are correct out of the total predictions made. It reflects the model's success across both positive and negative classes. A high accuracy rate may suggest that the model performs well across all classes. However, accuracy can be misleading, particularly when class distribution is imbalanced. For example, in a dataset where negative examples dominate, a model could achieve high accuracy by simply predicting negative for all cases. This overlooks the importance and challenges of the positive class, failing to accurately represent the model's ability to address real-world problems. Therefore, to evaluate a model's performance more fairly and comprehensively, it is crucial to consider accuracy alongside other metrics such as precision, recall, and F1-score [26]. The formula for the accuracy metric is shown in Equation (1).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

2.9.3 Precision

Precision is a performance metric that indicates how accurately a classification model identifies truly positive cases as positive. A high precision rate suggests that the model is selective and accurate in its positive classifications, meaning it correctly identifies positive instances with minimal false positive errors. Precision is particularly important in scenarios where false positives can lead to significant issues. For example, in disease detection, a diagnostic test with high precision reduces unnecessary anxiety, stress, and treatment costs caused by false positive results. Similarly, in spam filters, high precision helps prevent legitimate emails from being mistakenly marked as spam [26]. The formula for the precision metric is shown in Equation (2).

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

2.9.4 Recall

Recall is a statistical metric that measures how effectively a classification model captures truly positive cases. A high recall value indicates that the model is unlikely to miss positive instances. This is particularly crucial in applications where overlooking critical cases could lead to severe consequences. In fields like medical diagnostics, where missing a disease could have serious implications for a patient, recall is vital. A diagnostic system with high recall ensures that the disease is detected in as many cases as possible, increasing the chances of early intervention and improving treatment outcomes. In situations where false negatives carry significant costs, recall deserves special attention when evaluating model performance. Therefore, for a balanced performance assessment, recall should be considered alongside precision and other metrics [26]. The formula for the recall metric is shown in Equation (3).

$$\text{Recall} = \frac{\text{TP}}{TP + FN} \tag{3}$$

2.9.5 F1-Score

F1 Score is a metric that combines two critical performance measures of a classification model precision and recall into a balanced indicator. It represents the weighted harmony of these two metrics and is particularly useful when both are considered equally important. The F1 Score is calculated as the harmonic mean of precision and recall, requiring both values to be high; if either is low, it significantly lowers the F1 Score. This approach penalizes both false positive and false negative predictions in a balanced way. The F1 Score is more meaningful than accuracy alone in cases of imbalanced datasets or when correct classification of different classes carries varying costs or importance. For instance, in disease detection, high recall is crucial, while in spam email filtering, high precision is preferred. The F1 Score helps assess how well a model performs across both scenarios by finding a balance between these metrics. It is especially reliable in situations where one class is underrepresented or when incorrect predictions are costly, making it a trusted performance indicator [26]. The formula for the F1 Score metric is shown in Equation (4).

$$\text{F1} - \text{Score} = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4}$$

3 Results and discussion

For empirical validation of our mask detection frameworks, we selected the comprehensive "Mask Detection" dataset available through Kaggle's repository system. This curated collection offered a robust distribution

across relevant categories: the training partition contained 3,966 properly masked facial images, 3,668 unmasked subjects, and 1,674 instances of improper mask usage. The testing corpus comprised 529, 489, and 219 images respectively across these same categories, while the validation set included 794 properly masked, 734 unmasked, and 335 improperly masked facial captures. Prior to algorithmic processing, all imagery underwent standardization to 224×224 pixel dimensions a resolution determined through preliminary experimentation to balance detail preservation against computational demands. Following this preprocessing stage, we independently trained ResNet101, MobileNet, and Xception architectures before integrating their predictive capacities via Weighted Voting mechanisms. This methodological foundation facilitated the development of three distinct classification paradigms, each representing an evolutionary advancement over its predecessor: The initial Mask Ensemble Learning configuration established our baseline approach, demonstrating the fundamental efficacy of architectural integration without additional optimization techniques. When subjected to rigorous evaluation against our test corpus, this foundational model achieved 96.2% classification accuracy alongside precision, recall, and F1-score measurements of 96.2%, 94.2%, and 95.2% respectively. Building upon these promising results, our second experimental configuration the Attention Mask Ensemble Learning model incorporated sophisticated attention mechanisms designed to dynamically adjust feature importance during classification. This attentional refinement yielded modest but consistent performance improvements across all evaluation metrics: 96.6% accuracy, 96.5% precision, 94.5% recall, and an F1-score of 95.6%. The culmination of our experimental progression manifested in the Bayes Mask Ensemble Learning architecture, which leveraged Bayesian optimization strategies to systematically refine hyperparameter selections across constituent models. This probabilistic approach to parameter tuning resulted in substantial performance gains, with the model achieving 98.2% accuracy, 96.2% precision, 97.4% recall, and a remarkable F1-score of 97.8%. As illustrated in Table 1, comparative analysis across these three architectural variants reveals a clear progression in classification efficacy, with each successive refinement addressing specific limitations of its predecessors. Particularly noteworthy is the substantial improvement in recall metrics observed in the Bayes-optimized configuration, suggesting enhanced sensitivity to properly identifying masked states across diverse facial presentations.

Table 1. Performance comparison of the models

Models	Bayes Mask Ensemble Learning	Attention Mask Ensemble Learning	Mask Ensemble Model
Accuracy	0.982	0.966	0.962
Recall	0.974	0.945	0.942
Precision	0.982	0.965	0.962
F1 Score	0.978	0.956	0.952

In this study, the accuracy performance of the developed models Mask Ensemble Learning (96.2%), Attention Mask Ensemble Learning (96.6%), and Bayes Mask Ensemble Learning (98.2%) was compared with similar studies in the literature. Sethi et al. (2021) achieved 95.2%, Himeur et al. (2023) 96.8%, Kaur et al. (2022) 94.5%, and Teboulbi et al. (2021) 97.3% accuracy rates. The Bayes Mask Ensemble Learning model stands out with the highest accuracy rate of 98.2% compared to all studies in the literature, while the other two models also delivered competitive results. This comparison demonstrates that the proposed approach offers superior performance in the field of mask detection. A detailed comparison is provided in Table 2.

Table 2. Comparison of proposed models with literature models based on accuracy

Models	Accuracy
Deep Learning-Based Model (Sethi et al., 2021)	95.20%
VGG-16 (Himeur et al., 2023)	96.80%
CNN Model (Kaur et al., 2022)	94.50%
AI-Based System (Teboulbi et al., 2021)	97.30%
Mask Ensemble Learning (Proposed Model 1)	96.20%
Attention Mask Ensemble Learning (Proposed Model 2)	96.60%
Bayes Mask Ensemble Learning (Proposed Model 3)	98.20%

This phase of the investigation focused on two ensemble architectures: the Mask Ensemble Learning model and its enhanced counterpart, the Attention Mask Ensemble Learning model. Both frameworks were fundamentally constructed by amalgamating the predictive outputs of three core deep learning algorithms ResNet101, MobileNet, and Xception through the established Weighted Voting methodology.

In its more straightforward configuration, the Mask Ensemble Learning model directly aggregated the predictions from these constituent algorithms. This initial approach yielded a performance benchmark of 0.962 accuracy when evaluated on the designated test dataset. Seeking to refine this outcome, the Attention Mask Ensemble Learning model introduced an additional layer of sophistication. Specifically, dense layer-based attention mechanisms were integrated into the Weighted Voting process. This strategic inclusion was designed to optimize the distribution of influence among the combined models, and indeed, it culminated in an improved accuracy of 0.966 on the same test dataset. It is important to note that, for both these ensemble configurations, the training regimen proceeded using their default hyperparameter settings; the Bayes Search optimization technique was not applied at this stage. The evolving validation accuracy trends for these models, which also chart the performance trajectories of the individual ResNet101, MobileNet, and Xception algorithms across 150 training epochs, are visually presented in Figure 10. This graphical representation offers a clear illustration of the incremental benefit conferred by the attention mechanisms, even in the absence of explicit hyperparameter optimization.

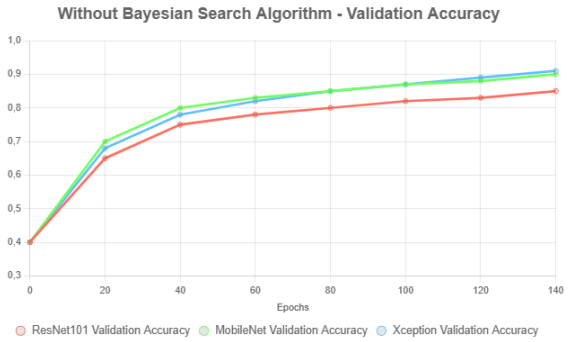

Figure 10. Comparative validation accuracy of deep learning models without Bayesian optimization

Our investigation culminated in the development of a sophisticated classification architecture leveraging Bayesian optimization techniques for hyperparameter tuning across three prominent convolutional networks: ResNet101, MobileNet, and Xception. Rather than accepting default configurations, we employed Bayes Search algorithms to systematically explore the parameter space before integrating these refined models through carefully calibrated weighted voting procedures. The convergence dynamics illustrated in Figure 11 provide compelling evidence for the efficacy of this approach. Examination of the validation accuracy trajectories across 150 training epochs reveals two critical phenomena: first, the Bayesian optimized networks demonstrate markedly accelerated learning during early training phases; second, they exhibit substantially reduced oscillatory behavior during later epochs, suggesting more robust generalization characteristics. This stability differential becomes particularly pronounced after approximately 100 epochs of training. Such convergence improvements directly translate to quantifiable performance advantages. When evaluated against our carefully constructed test dataset, the resulting Bayes Mask Ensemble Learning framework achieved classification accuracy of 0.982 a figure that not only surpasses the alternative approaches explored within our experimental framework but also exceeds previously published benchmarks in the facial mask detection domain. This empirical superiority underscores the considerable potential of probabilistic hyperparameter optimization in addressing classification challenges within computer vision applications.

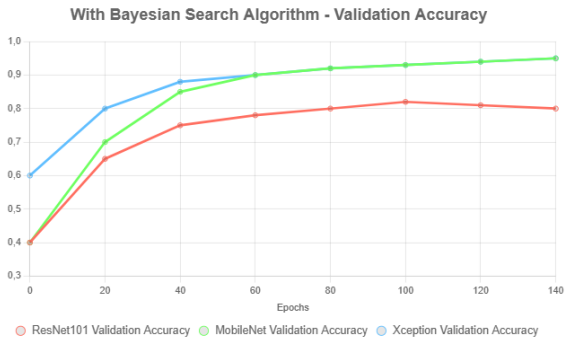

Figure 11. Comparative validation accuracy of deep learning models with Bayesian optimization

4 Conclusion

The experiments conducted in this study demonstrate that the proposed Mask Ensemble Learning, Attention Mask Ensemble Learning, and Bayes Mask Ensemble Learning models exhibit high performance in mask detection. The Mask Ensemble Learning model, using a basic ensemble approach, achieved a 96.2% accuracy rate, establishing a reliable foundation for mask detection. The Attention Mask Ensemble Learning model, incorporating attention mechanisms, further improved performance with a 96.6% accuracy rate, highlighting the contribution of attention mechanisms in optimizing feature weighting. The highest performance was delivered by the Bayes Mask Ensemble Learning model, which optimized hyperparameters using Bayes Search, achieving a 98.2% accuracy rate and surpassing other methods in the literature. These results indicate that the proposed system is highly competitive in mask detection applications. However, further improvements could be pursued to achieve even higher accuracy rates. For instance, incorporating more advanced attention mechanisms or alternative optimization algorithms could enhance model performance. Additionally, increasing the diversity and size of the dataset could improve the model's generalization ability. Integrating more complex deep learning architectures or conducting experiments with longer training durations and broader hyperparameter ranges could also positively impact performance.

Beyond these improvements, closely monitoring advancements in mask detection technologies and periodically reassessing the system are critical for ensuring its long-term effectiveness and sustainability. Moreover, practical enhancements, such as optimizing the system for low-power hardware and reducing processing times, could be considered to improve performance in real-time applications.

In conclusion, the 98.2% accuracy rate of the Bayes Mask Ensemble Learning model clearly demonstrates the system's potential in mask detection while also revealing areas for improvement. Detailed analyses and optimization efforts could further enhance the system's performance, positioning it as even more competitive in mask detection applications. This study underscores the importance of AI-based solutions in combating infectious diseases and provides a significant reference point for improving the effectiveness of automated systems in monitoring mask compliance in public spaces. The findings offer a valuable foundation for researchers, developers, and decision-makers in the development and implementation of such technologies.

Conflict of Interest

The authors declare that there are no conflicts of interest regarding this study.

Similarity Rate (iThenticate): 12%

References

[1] F. A. Muhammed Ali and M. S. Al-Tamimi, Face mask detection methods and techniques: A review. International Journal of Nonlinear Analysis and

Applications, 13 (1), 3811-3823, 2022. http://dx.doi.org/10.22075/ijnaa.2022.6166.

[2] J. G. Chowdary, N. S. Punn, S. K. Sonbhadra and S. Agarwal, Face mask detection using transfer learning of InceptionV3. Big Data Analytics: 8th International Conference (BDA 2020), pp. 81-90, Sonepat, India, 15-18 December 2020. https://doi.org/10.1007/978-3-030-66665-1_6.

[3] H. Adusumalli, D. Kalyani, R. K. Sri, M. Pratapteja and P. P. Rao, Face mask detection using OpenCV. Third International Conference on Smart Communication Technologies and Virtual Mobile Networks (ICICV), pp. 1304-1309, IEEE, Tirunelveli, India, 11-12 February 2021. https://doi.org/10.1109/ICICV50876.2021.9388375.

[4] S. Sethi, M. Kathuria and T. Kaushik, Face mask detection using deep learning: An approach to reduce risk of Coronavirus spread. Journal of Biomedical Informatics, 120, 103848, 1-10, 2021. https://doi.org/10.1016/j.jbi.2021.103848.

[5] K. M. Hosny, N. A. Ibrahim, E. R. Mohamed and H. M. Hamza, Artificial intelligence-based masked face detection: A survey. Intelligent Systems with Applications, 22, 200391, 1-22, 2024. https://doi.org/10.1016/j.iswa.2024.200391.

[6] Y. Himeur, S. Al-Maadeed, I. Varlamis, N. Al-Maadeed, K. Abualsaud and A. Mohamed, Face mask detection in smart cities using deep and transfer learning: Lessons learned from the COVID-19 pandemic. Systems, 11 (2), 107, 1-28, 2023. https://doi.org/10.3390/systems11020107.

[7] G. Kaur, R. Sinha, P. K. Tiwari, S. K. Yadav, P. Pandey, R. Raj, A. Vashisth and M. Rakhra, Face mask recognition system using CNN model. Neuroscience Informatics, 2 (3), 100035, 1-9, 2022. https://doi.org/10.1016/j.neuri.2021.100035.

[8] S. Teboulbi, S. Messaoud, M. A. Hajjaji and A. Mtibaa, Real-time implementation of AI-based face mask detection and social distancing measuring system for COVID-19 prevention. Scientific Programming, 2021 (1), 8340779, 1-14, 2021. https://doi.org/10.1155/2021/8340779.

[9] M. Loey, G. Manogaran, M. H. N. Taha and N. E. M. Khalifa, Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. Sustainable Cities and Society, 65, 102600, 1-9, 2021. https://doi.org/10.1016/j.scs.2020.102600.

[10] A. Chavda, J. Dsouza, S. Badgujar and A. Damani, Multi-stage CNN architecture for face mask detection. 6th International Conference for Convergence in Technology (I2CT), pp. 1-8, IEEE, Maharashtra, India, 2-3 April 2021. https://doi.org/10.1109/I2CT51068.2021.9418207.

[11] B. Qin and D. Li, Identifying facemask-wearing condition using image super-resolution with classification network to prevent COVID-19. Sensors, 20 (18), 5236, 1-16, 2020. https://doi.org/10.3390/s20185236.

[12] S. Yadav, Deep learning based safe social distancing and face mask detection in public areas for COVID-19 safety guidelines adherence. International Journal for Research in Applied Science and Engineering Technology, 8 (7), 1368-1375, 2020. https://doi.org/10.22214/ijraset.2020.30560.

[13] S. Shiekh Burhan, Face Mask Dataset. Kaggle, 2020. https://www.kaggle.com/datasets/shiekhburhan/face-mask-dataset, Accessed 4 March 2025.

[14] Y. Tong, W. Lu, Q. Q. Deng, C. Chen and Y. Shen, Automated identification of retinopathy of prematurity by image-based deep learning. Eye and Vision, 7, 40, 1-12, 2020. https://doi.org/10.1186/s40662-020-00206-2.

[15] Z. Wu, C. Shen and A. Van Den Hengel, Wider or deeper: Revisiting the ResNet model for visual recognition. Pattern Recognition, 90, 119-133, 2019. https://doi.org/10.1016/j.patcog.2019.01.006.

[16] N. S. Sanjay and A. Ahmadinia, MobileNet-Tiny: A deep neural network-based real-time object detection for Raspberry Pi. 2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 647-652, IEEE, Boca Raton, FL, USA, 16-19 December 2019. https://doi.org/10.1109/ICMLA.2019.00118.

[17] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand and H. Adam, MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint, arXiv:1704.04861, 1-9, 2017. https://doi.org/10.48550/arXiv.1704.04861.

[18] Y. Liu, L. Zhang, Z. Hao, Z. Yang, S. Wang, X. Zhou and Q. Chang, An Xception model based on residual attention mechanism for the classification of benign and malignant gastric ulcers. Scientific Reports, 12 (1), 15365, 1-13, 2022. https://doi.org/10.1038/s41598-022-19639-x.

[19] F. Chollet, Xception: Deep learning with depthwise separable convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1251-1258, IEEE, Honolulu, HI, USA, 21-26 July 2017. https://doi.org/10.1109/CVPR.2017.195.

[20] V. Nguyen, Bayesian optimization for accelerating hyper-parameter tuning. 2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), pp. 302-305, IEEE, Cagliari, Italy, 3-5 June 2019. https://doi.org/10.1109/AIKE.2019.00060.

[21] M. Feurer and F. Hutter, Hyperparameter optimization. in Automated Machine Learning: Methods, Systems, Challenges. F. Hutter, L. Kotthoff, J. Vanschoren (Eds.), Springer, pp. 3-33, Cham, 2019. https://doi.org/10.1007/978-3-030-05318-5_1.

[22] A. Doğan and D. Birant, A Weighted majority voting ensemble approach for classification. 2019 4th International Conference on Computer Science and Engineering (UBMK), pp. 1-6, IEEE, Samsun, Turkey, 11-15 September 2019. https://doi.org/10.1109/UBMK.2019.8907028.

[23] J. M. Banda, R. A. Angryk and P. C. H. Martens, Steps toward a large-scale solar image data analysis to differentiate solar phenomena. Solar Physics, 288, 435-462, 2013. https://doi.org/10.1007/s11207-013-0304-x.

[24] C. Hark, Sahte haber tespiti için derin bağlamsal kelime gömülmeleri ve sinirsel ağların performans değerlendirmesi. Fırat Üniversitesi Mühendislik Bilimleri Dergisi, 34 (2), 733-742, 2022. https://doi.org/10.35234/fumbd.1126688.

[25] A. Kulkarni, D. Chong and F. A. Batarseh, Foundations of data imbalance and solutions for a data democracy. in Data Democracy. F. A. Batarseh, R. Freeman (Eds.), Academic Press, pp. 83-106, London, 2020. https://doi.org/10.1016/B978-0-12-818366-3.00005-8.

[26] D. M. Powers, Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv preprint, arXiv:2010.16061, 1-23, 2020. https://doi.org/10.48550/arXiv.2010.16061.