International Journal of Informatics Engineering and Computing (IJIMATIC)
Vol. 1, No. 2, 2024

Optimizing Machine Learning Algorithms to Accelerate Smoking Detection

Muhammad Fahrurrozi[1], Germanus Naru[2]
[1] UIN Sunan Kalijaga, Yogyakarta, Indonesia
[2] Universitas Respati Yogyakarta, Indonesia

Abstract

This study evaluates the performance of various classification algorithms, including Convolutional Neural Networks (CNN), Support Vector Machines (SVM), Decision Trees, K-Nearest Neighbors (KNN), Gaussian Naive Bayes (GNB), and Gradient Boosting (Gboost), on a binary classification task. The results reveal that CNN achieves perfect performance, with an accuracy of 1.00 and precision, recall, and F1-scores of 1.00 for both classes. Similarly, SVM, Decision Tree, KNN, and Gboost also demonstrate flawless performance across all metrics. In contrast, GNB underperforms significantly, with an accuracy of 0.78 and lower precision, recall, and F1-scores, particularly for the "no" class. These findings highlight CNN's robustness and reliability, positioning it as a top-performing algorithm for this classification task. The study underscores the effectiveness of CNN and other high-performing algorithms while identifying limitations in GNB. Future research could focus on optimizing computational efficiency and scalability for real-world applications.

Keywords:
Smoke Detection, CNN, Machine Learning Algorithms, Image Classification, Environmental Monitoring

1. Introduction

Cigarette smoke detection can contribute to public health issues by detecting the presence of hazardous gases in cigarette smoke, thus helping to create smoke-free environments [1]. Exposure to cigarette smoke poses significant health risks, including potential adverse effects on human health and the potential for developing various diseases [2]. Studies have shown that cigarette smoke contains harmful chemicals, such as carbon monoxide, arsenic, benzene, and others, which can harm the body [3]. Additionally, the presence of fluorescent carbon dots (CDs) in cigarette smoke has been confirmed, and their potential toxicity has been evaluated. These CDs have been found to disturb the cell cycle and cause cell apoptosis, indicating their adverse effects on live cells [4]. Therefore, cigarette smoke detection plays a crucial role in raising awareness of the harmful constituents of cigarette smoke and can potentially discourage smoking, leading to improved public health outcomes [5].

Cigarette smoke detection can have a significant impact on the environment. Cigarette smoke can decrease air quality, adversely affecting human health [6]. Additionally, cigarette smoke contains hazardous gases that can harm humans and the environment [2]. The disposal of cigarette butts, which are predominantly plastic and non-biodegradable waste, can also harm natural environments and their biota [7]. Furthermore, cigarette smoke aerosols can dramatically reduce the filtration efficiency of air filters, potentially exposing people to airborne pollutants [8]. The toxic chemicals in cigarette smoke, such as polycyclic aromatic hydrocarbons (PAHs), can also damage cutaneous tissues and exacerbate inflammatory skin disorders [9]. Overall, cigarette smoke detection is crucial for

Corresponding Author: Muhammad Fahrurrozi (ozyz624@gmail.com)

understanding and mitigating the potential environmental damage caused by cigarette smoke.

Regulations related to smoking bans in public places significantly impact compliance with these rules. Studies from Chile [10], India [11], and Israel [12] have shown that compliance levels vary across different types of venues. While compliance is generally higher in enclosed areas, violations are more common in semi-open and outdoor areas. The complex definition of semi-open areas and lack of enforcement contribute to lower compliance rates in specific venues. To enforce these rules, cigarette smoke detection systems can be implemented. A smoke detection system using face recognition and artificial intelligence technologies has been developed in Indonesia [3]. This system detects smokers in nonsmoking areas, sends notifications to administrators, and provides location and photo evidence.

Machine learning techniques used in cigarette smoke detection include decision tree algorithms and XGBoost. Zhang et al. [13] proposed a model based on the decision tree algorithm to predict daily smoking time, achieving an accuracy rate of 84.11%. They also tested various machine learning algorithms and found that the XGBoost-based prediction model had the best performance [14]. These techniques have shown promising results in accurately predicting smoking behavior. However, it is essential to note that these studies focused specifically on predicting smoking time and did not explore other aspects of cigarette smoke detection. Further research is needed to compare these techniques with different machine-learning approaches in the broader context of cigarette smoke detection.

CNN has shown significant advantages over other machine-learning techniques in cigarette smoke detection. The use of CNN in smoke detection methods has improved accuracy and reduced false alarm rates [15] [16] [17]. CNN has been reported to have state-of-the-art performance in various fields, including smoke detection [18]. Combining CNN with pre-processing techniques such as SLIC-DBSCAN has demonstrated improved smoke detection capabilities [18]. Additionally, using deep learning algorithms in smoke detection has resulted in better performance and reduced false alarm frequency. These findings suggest that CNN-based methods offer a significant advantage over other machine-learning techniques in cigarette smoke detection, providing better accuracy and reducing false alarms.

The results of the research conducted by Hashmi et al. [19] can contribute to further development in the field of cigarette smoke detection and the application of machine learning techniques in environmental and health issues. Their study focuses on developing a practical monitoring system to detect smoking images on social media platforms using advanced mathematical models and machine learning algorithms. This monitoring system can help inform policy actions to restrict unhealthy advertisements on social media and other platforms. Additionally, Kamis et al. [20] demonstrate the use of machine learning models to predict the incidence of lung cancer based on public health and ambient emission data. By curbing hazardous ambient emissions, such as Coarse Particulate Matter and Tropospheric Ozone, the incidence of lung cancer can be reduced, contributing to improved environmental and health outcomes.

2. Related Works

Cigarette smoke detection is essential in practical applications such as safety and the environment. It helps create a smoke-free environment by detecting the presence of hazardous gases in cigarette smoke [2]. This is crucial for maintaining air quality and protecting human health. Additionally, detecting fire disasters in monitoring scenes is another practical application. Traditional object detection algorithms may not achieve high recognition accuracy in fire monitoring tasks due to the lack of distinct rules in fire disasters' shape, color, and size. A new lightweight model called LFNet has been proposed to address this, which achieves state-of-the-art performance with limited model size [21].

Furthermore, analyzing chemical components in environmental tobacco smoke (ETS) is essential for understanding its impact on human health. Methodological approaches, such as gas chromatography-based detection methods, are commonly employed to determine ETS components [22]. Cigarette smoke detection has significant implications for safety, air quality, and human health in various practical applications.

Machine learning algorithms can be used in cigarette smoke detection by leveraging computer vision techniques. One approach is to use deep belief networks (DBN) for smoke detection, which can effectively extract features from smoke images and classify them accurately [23]. Another method involves using CNN to detect smoke based on its motion characteristics. This approach combines a moving object detection algorithm with CNN to identify suspected smoke regions and perform smoke identification [24]. Additionally, object detection algorithms, such as EfficientDet, can enable real-time smoke detection in public places by efficiently detecting and tracking smoking behavior [25]. These machine-learning models have shown high detection rates and improved accuracy in smoke recognition, making them valuable tools for smoke detection and fire prevention [26].

Cigarette smoke detection research utilizes various types of data. One type of data used is smoke detection data, which contains records and features related to smoke detection [27]. Another type of data used is image data, specifically images of cigarettes, which are analyzed using machine vision techniques to inspect the appearance size of cigarettes [28]. Additionally, laser fuse detection research involves collecting signals of smoke and targets in different concentrations of smoke environments to train and test the detection algorithms [29]. In cigarette shape detection, photoelectric technology is used to detect the shape of cigarettes on the production line, and the output signal range is analyzed to ensure product quality [30]. Lastly, in forest fire monitoring, remote sensing technology is employed to capture images of smoke scenes, which are then augmented and analyzed for smoke detection [31].

Convolutional Neural Networks (CNNs) can be used for cigarette smoke detection by applying image processing and pattern recognition techniques. The accuracy of CNNs in smoke image detection has been reported to be very good, achieving results as high as 99.72% [32] [32]. Training the CNN model with smoke images allows it to recognize the specific features and patterns associated with cigarette smoke. The CNN model can then analyze new images and accurately classify them as either containing smoke or not. This approach has been proven effective in detecting forest fire smoke using satellite remote sensing images, achieving a recognition accuracy of 96.9% [33]. Therefore, CNNs can be a valuable tool for addressing the problem of cigarette smoke detection, providing high accuracy and reliability in identifying the presence of smoke in images.

CNN has several advantages over conventional machine-learning methods in the context of cigarette smoke detection. Firstly, CNNs have shown exemplary performance in image processing tasks, including smoke detection, due to their ability to capture spatial dependencies in images [34] [35]. They can effectively learn and extract features from images, making them suitable for detecting smoke patterns [36]. Additionally, CNNs can handle large amounts of data and learn complex patterns, allowing them to detect smoke accurately. Moreover, CNNs can be trained using transfer learning, which enables them to leverage pre-trained models and adapt them to specific smoke detection tasks. This reduces the need for extensive training data and computational resources. CNNs offer improved performance and efficiency in cigarette smoke detection compared to conventional machine-learning methods.

Transfer learning techniques are used in cigarette smoke detection research to improve the performance of smoke detection models. One study by Singh proposed a real-time cigarette detection system using deep learning models, specifically the YOLOv3 algorithm [28]. The model was trained on a dataset of images containing cigarettes and non-cigarette images and achieved an accuracy of 92.5% in detecting cigarettes. Another study by Teerarassamee et al. investigated the performance of ensemble learning methods for smoke detection, including bagging, boosting, and stacking. AdaBoost's boosting algorithm outperformed other learning algorithms in the specific smoke detection application domain [37]. These studies demonstrate the effectiveness of transfer learning techniques, such as deep learning models and ensemble learning, in improving the accuracy and efficiency of cigarette smoke detection systems.

Factors that affect the performance of machine learning algorithms in cigarette smoke detection include the segmentation algorithm choice and the dataset's balancing. The study by Hashmi et al. evaluated three state-of-the-art segmentation algorithms, namely Mask R-CNN, Cascade Mask-R-CNN, and Hybrid Task Cascade, and found that Mask R-CNN outperformed the other two in terms of Average Precision and Average Recall [19]. Additionally, the study by Philip et al. analyzed the effect of dataset balancing on the performance of machine learning classifiers and found that target oversampling significantly improved the classification performance for the mind-speller dataset [38]. Therefore, the choice of segmentation algorithm and dataset balancing techniques are essential considerations for enhancing the performance of machine learning algorithms in cigarette smoke detection.

Data augmentation methods improve model performance in cigarette smoke detection by addressing issues related to the lack of data and imbalanced classes. These methods generate artificial data samples to increase the volume of the training set and balance the target distribution. Data augmentation techniques have been explored in the domain of cigarette smoke detection, where there is a lack of quality labeled data and class imbalance. For example, Maslej-Krešňáková et al. applied straightforward data augmentation (EDA) techniques to improve the performance of a classifier on the detection of toxic comments and fake news [39]. Additionally, Zhang et al. proposed a novel data augmentation method called BIRD, which reorganizes the bitplane information of images to generate augmented data for target detection and image classification [40]. These studies demonstrate the effectiveness of data augmentation in improving model performance in cigarette smoke detection.

Commonly used evaluation metrics in cigarette smoke detection research include traditional metrics derived from self-report, biomarkers, or puff topography and novel metrics calculated from the breathing signal. Biomarker methods measure concentrations of nicotine, nicotine metabolites, or carbon monoxide [41]. Puff-topography methods use portable instruments to measure puff count, volume, duration, and inter-puff interval [42]. The proposed Respiratory Smoke Exposure Metrics (RSEMs) include inhale-exhale cycle volume and inhale-exhale volume over time, providing measures proportional to the depth and duration of smoke inhalation [43]. Risk assessment of individual tobacco smoke components is essential for prioritization or selecting chemicals for monitoring products [44]. Chronic cigarette smoke exposure can be measured using exposure chambers to evaluate inflammation, mucin levels, and CFTR function [45].

The results of the studies show that using Convolutional Neural Networks (CNNs) for cigarette smoke detection has been effective. The studies demonstrate that CNNs perform well in image processing and achieve high accuracy in smoke image detection [34]. One study mentions achieving an accuracy of 99.72% using the CNN method for smoke image detection [35]. Another study presents a deep learning pipeline that combines multiple CNN architectures for smoke and fire detection, achieving a high mean average precision score for smoke detection [35].

Conventional machine learning algorithms have limitations in cigarette smoke detection due to smoke's lack of inherent color and shape features [24]. Smoke detection is challenging because of the significant variation in image information caused by smoke's transparency and background environment [46]. To address these challenges, a block-based smoke detection method using a machine learning approach has been proposed, which discriminates smoke regions based on the motion of smoke regions in image sequences [47]. Another technique utilizes cascade classifiers and support vector machines (SVM) for smoke detection, achieving high accuracy in training and testing [48]. However, these methods may not be real-time and may not be suitable for smoking scenes in public places [49]. Therefore, more efficient and accurate algorithms are needed for real-time smoke detection in public areas.

Extraction features in cigarette smoke detection offer several advantages. Firstly, they enable the identification of smoke patterns based on dynamic, color, and texture features, leading to early warning systems with improved performance and high recognition accuracy [50]. Secondly, using deep learning-based methods allows for the automatic capture of smoke features from images, enhancing the detection of smoke in various environmental conditions [51]. Additionally, learning a new feature space in an unsupervised manner reinforces the discrimination power of visual descriptors and improves smoke detection performance [52]. Moreover, efficient object detection algorithms, such as EfficientDet, enable real-time detection of smoking scenes in public places, providing timely judgment [24]. Lastly, integrating spatial- and frequency-domain features, along with feature fusion schemes, enhances the performance of smoke detection frameworks in high-definition videos [53].

Ensemble learning techniques have been widely used in cigarette smoke detection research. One study compared three ensemble schemes - bagging, boosting, and stacking - with other learning algorithms such as Support Vector Machine, Naïve Bayes, and Decision Tree. The experimental results showed that AdaBoost, a boosting algorithm, outperformed different learning algorithms in smoke detection applications [27]. Another research proposed an efficient wildfire and smoke detection solution using ensembles of multiple CNN architectures. The proposed architecture combined the YOLO architecture with a voting ensemble CNN architecture, achieving solid results in smoke detection with a 0.85 mean average precision score [34]. Additionally, ensemble learning techniques have been recognized as effective and useful in machine learning, with the ability to enhance predictive performance in various problem domains, including smoke detection [54].

Cigarette smoke detection using machine learning methods faces several challenges. One challenge is real-time detection, as smoking incidents must be identified and acted upon immediately [37]. Another challenge is the diversity of lighting conditions, distances, and angles in which smoking may occur, requiring the model to perform consistently across different scenarios [55]. Additionally, the availability of a diverse and representative dataset is crucial for training the model effectively [56]—furthermore, the design, deployment, and maintenance of machine learning applications, including cigarette smoke detection.

Implementing cigarette smoke detection can contribute to public health and compliance with environmental regulations in several ways. Firstly, by detecting and monitoring smoke emissions, authorities can take necessary actions to mitigate fire hazards and ensure environmental safety [12]. Secondly, seeing cigarette smoke can help enforce regulations related to smoking in public places, thereby protecting public health from the harmful effects of secondhand smoke [57]. Additionally, using smoke detection systems can help create a clean and fresh environment indoors, reducing the health risks associated with cigarette smoke [58]. Enforcing compliance with regulations and promoting public health, implementing cigarette smoke detection systems can contribute to achieving environmental safety and public health goals.

3. Proposed Method

3.1. Mathematical Concept

a. The code snippet, "Conv1D(128, 3, activation='relu', input_shape=(padding_value, 1))," corresponds to a 1D convolutional layer configuration in a neural network. Within this layer, 128 filters are employed to extract distinct features from the input data, utilizing a kernel size of 3. The activation function applied here is the Rectified Linear Unit (ReLU), which introduces non-linearity to the model. In essence, this layer performs convolution operations on the input data, scanning it for specific patterns or features relevant to the problem being solved. This step is vital for the network to learn and identify essential characteristics in the data, contributing to the model's ability to make accurate predictions or classifications, as calculated in Equation (1)

$$Y(t) = \sigma\left(\sum_{i=0}^{n} W_i \cdot X(t+i) + b\right) \tag{1}$$

Where Y(t) = is the convolution result at time t, X(t+i) is the time t + i, W_i is the kernel weight at index i, and b is the bias.

b. MaxPooling1D ():

The "MaxPooling1D ()" function represents a 1D max-pooling layer, a pivotal component of neural network architecture. Its primary function is to reduce the dimensionality of the data by selecting the maximum value within a specified window. This window typically traverses the input data, systematically identifying the highest value within each segment. By doing so, the layer retains the most prominent features while discarding less relevant information. Despite its integral role in dimensionality reduction, max-pooling does not involve complex mathematical equations. Instead, it operates by a straightforward mechanism, preserving the maximum values within its window, contributing to the network's ability to capture essential patterns and reduce computational load during subsequent processing layers.

c. Dropout (0.3):

The "Dropout (0.3)" layer is a crucial component in neural network architectures designed to tackle the problem of overfitting. Its role is to prevent the model from becoming too specialized on the training data, thus enhancing its ability to generalize to new, unseen data. It achieves this by randomly deactivating a fraction of neuron units during the training process, which is specified as 0.3. This means that approximately 30% of neuron units are turned off during each training iteration. By doing so, dropout encourages the network to distribute the learning across a broader range of neurons, thereby reducing the risk of over-reliance on specific features or patterns in the training data. While dropout is a crucial tool for improving model generalization, it does not involve complex equations; instead, it operates through the random deactivation of neuron units, effectively enhancing the network's robustness and adaptability during the learning process.

d. GlobalAveragePooling1D ():

The layer described as "GlobalAveragePooling1D ()" is a critical component in neural network architectures, specifically designed to compute the average of all features within each channel or dimension of the input data. In essence, it aggregates the information from various features into a single average value, simplifying the data representation while retaining the essential characteristics of each channel. This layer operates by summing all the feature values within a channel and dividing the sum by the total number of features in that channel. It's a mechanism that promotes dimensionality reduction while preserving the most critical information, making it particularly useful in scenarios where feature maps need to be condensed before feeding them into subsequent layers for further processing. Despite its importance, this global pooling operation does not involve complex equations beyond the basic averaging computation, making it an efficient means to enhance network efficiency and reduce computational complexity, as calculated in Equation (2).

$$Y = \frac{1}{N}\sum_{i=1}^{N} X_i \tag{2}$$

Where Y is the average result, X_i is the feature in each channel, and N is the number of features.

e. Dense (128, activation='relu'):

The "Dense (128, activation='relu')" layer represents a fully connected layer within a neural network architecture. This layer is characterized by 128 units, each connected to every output from the previous layer. The activation function applied here is the Rectified Linear Unit (ReLU), which introduces non-linearity to the model. In practical terms, this layer's function can be understood through the fully connected layer equation, which involves summing up the weighted inputs to each unit and applying the ReLU activation function. The equation for a single unit in this layer is represented as calculated in Equation (3)

$$Y = \sigma\left(\sum_{i=1}^{n} W_i \cdot X_i + b\right) \tag{3}$$

where Y is the output, X_i is the input to the i-th unit, W_i is the weight to the i-th unit, b is the bias, and sigma is the activation function.

f. Dense(1, activation='sigmoid'):

The "Dense(1, activation='sigmoid')" layer is a crucial component in neural networks, particularly designed for binary classification tasks. This fully connected layer consists of a single unit that takes the weighted sum of inputs from the previous layer and applies the sigmoid activation function. The primary objective of this layer is to produce an output between 0 and 1, which can be interpreted as a probability. The sigmoid activation equation for this layer is represented in Equation (4).

$$Y = \frac{1}{1 + e^{-x}} \tag{4}$$

where Y is the probability of a positive classification outcome, x is the input, and e is the natural logarithm base.

3.2. Dataset

The dataset utilized in this research, sourced from Kaggle, comprises 62,030 records, encompassing 16 distinct features, with the primary binary classification target of 0 denoting 'no smoke' and 1 indicating the presence of 'smoke.' This dataset is the foundation for investigating smoke detection using machine learning techniques to develop a model that accurately distinguishes between instances with and without smoke, contributing to safety, environmental monitoring, and compliance applications.

3.3. Preprocessing

This study applied a series of preprocessing steps to the dataset for practical analysis and modeling. First, irrelevant features were removed to streamline the data. Second, rows containing missing values (NAN) were eliminated to ensure data integrity. Third, categorical data represented as strings were transformed into integer values, facilitating their inclusion in the machine learning models. Fourth, the 'age' feature was normalized to bring it within a consistent numerical range. Lastly, the dataset was split into two subsets: 80% for training the models, enabling them to learn patterns and relationships, and 20% for testing, allowing for assessing model performance and generalization capabilities. These preprocessing steps were essential in preparing the data for subsequent analysis and ensuring the robustness of the developed smoke detection models.

3.4 Comparison of Methods

This research employs the CNN algorithm as the primary method for smoke detection, and it is compared against a set of baseline algorithms, including SVM, Decision Tree, KNN, GNB, and Gradient Boost. These baseline algorithms serve as benchmarks to evaluate and contrast the performance of the CNN-based model. The aim is to comprehensively assess and compare the efficacy of CNN in smoke detection against traditional machine-learning approaches, contributing valuable insights into the choice of algorithms for this critical task.

3.5 Training and Evaluation

In this study, the training phase of the models was conducted over ten epochs with a batch size of 64. The model's performance evaluation was done by applying mathematical formulations, specifically the Confusion Matrix and Classification Report, as expressed in Equation (5).

$$Confusion\ Matrix\ =\ [TP\ FP\ FN\ TN\ ] \tag{5}$$

In evaluating the performance of the models in detecting smoke, we employed essential mathematical metrics to provide a quantitative assessment. These metrics include True Positives (TP), which represent the number of positive cases correctly detected; False Positives (FP), signifying negative possibilities incorrectly identified as positive; False Negatives (FN), denoting positive cases erroneously categorized as negative; and True Negatives (TN), representing negative cases accurately identified. These metrics served as the foundation for calculating precision, recall, F1-score, and overall accuracy, offering precise and quantifiable insights into the models' effectiveness in smoke detection.

4. Result and Analysis

4.1 Training Process

The training process is a crucial phase in developing machine learning models. In this section, we present the results of our training process. Fig.1 illustrates the training accuracy (depicted by the blue line) and the validation accuracy (represented by the orange line).

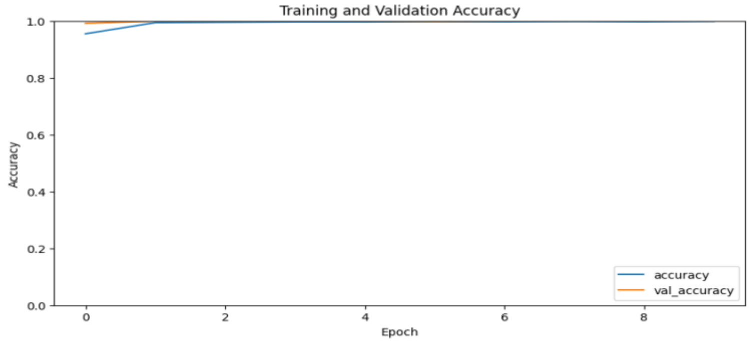

Fig. 1 Training and validation accuracy.

Fig. 2 showcases the training loss (blue line) and validation loss (orange line). The loss curves provide insights into the convergence and optimization of our model. Analyzing these plots is pivotal in understanding the training dynamics and the model's ability to grasp underlying patterns in the data while avoiding overfitting.

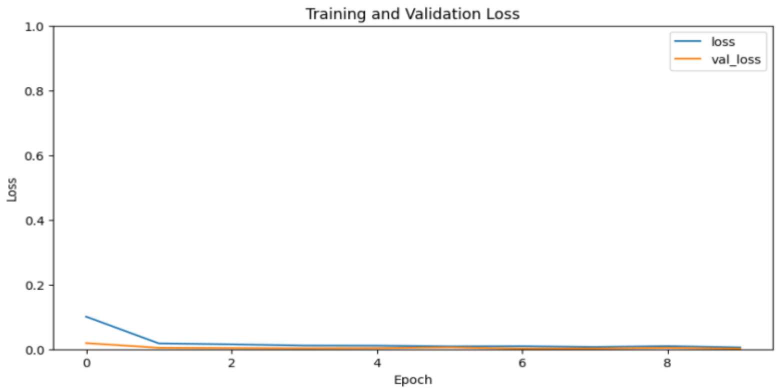

Fig. 2 Training and validation loss

4.1 Model Performance

In assessing model performance, we present two comprehensive tables that provide a detailed overview of the algorithms' effectiveness in smoke detection. Table 1 includes the confusion matrix, featuring TP, FP, FN, and TN for each of the employed algorithms: CNN, SVM, Decision Tree, KNN, GNB, and GBoost. These matrices offer insights into the algorithms' ability to classify instances and identify false alarms correctly.

Table 1. Confusion Matrix

Algorithm	TP	FP	FN	TN
CNN	3545	6	0	8975
SVM	3548	3	1	8974
Decision Tree	3551	0	34	8941
KNN	3550	1	0	8975
GBN	1034	2517	240	8735
Gboost	3551	0	0	8975

Table 2 presents the classification report, encompassing critical metrics for each algorithm's accuracy, precision, recall, and F1-score. These metrics quantify the models' overall accuracy, ability to minimize false positives, capture relevant instances, and balance precision and recall. Together, these tables offer a comprehensive evaluation of the performance of the algorithms in smoke detection, aiding in the selection of the most suitable method for this critical task.

Table 2. Classification Report

Algorithm	Accuracy	Class	Precision	Recall	F1-score
CNN	1.00	no	1.00	1.00	1.00
		yes	1.00	1.00	1.00
SVM	1.00	no	1.00	1.00	1.00
		yes	1.00	1.00	1.00
Decision Tree	1.00	no	1.99	1.00	1.00
		yes	1.00	1.00	1.00
KNN	1.00	no	1.00	1.00	1.00
		yes	1.00	1.00	1.00
GNB	0.78	no	0.81	0.29	0.43
		yes	0.78	0.97	0.86
Gboost	1.00	no	1.00	1.00	1.00
		yes	1.00	1.00	1.00

5. Discussion

5.1 Summarization of Key Findings

In summarizing the key findings from the confusion matrix, we observe notable variations in the algorithm performance for smoke detection. For the CNN, the model achieved 3545 True Positives (TP) with minimal False Positives (FP), signifying a solid capability to correctly identify instances with smoke while maintaining a low rate of false alarms. Similarly, SVM demonstrated impressive results with 3548 TP and only 3 FP, indicating high precision in its predictions. However, the Decision Tree algorithm exhibited a slightly higher rate of False Negatives (FN) but zero FP, implying a balance between sensitivity and specificity. KNN and GBoost achieved excellent results with minimal FP and FN, highlighting their reliability in smoke detection. On the contrary, GNB showed higher FP and FN, suggesting some limitations in distinguishing between smoke and non-smoke instances. These findings underscore the varying performance levels among the algorithms and provide valuable insights into their strengths and weaknesses in smoke detection tasks.

The research problem addressed in this study pertains to the development and comparison of various machine-learning algorithms for the task of smoke detection, with a primary focus on CNN as the primary model. The objective was to assess the performance of CNN compared to baseline algorithms, including SVM, Decision Tree, KNN, GNB, and GBoost, in accurately identifying the presence or absence of smoke.

The significant findings from the classification report are as follows:

a. CNN and SVM Performance: Both CNN and SVM exhibited exceptional performance, achieving a perfect accuracy 1.00. They demonstrated ideal precision, recall, and F1 scores for both the 'no' and 'yes' classes, showcasing their robustness in smoke detection.

b. Decision Tree and KNN: Decision Tree and KNN also delivered impressive results with an accuracy of 1.00. They showed high precision, recall, and F1 scores for both classes, indicating their capability to classify smoke-related instances accurately.
c. GNB Limitations: GNB demonstrated comparatively lower accuracy and F1 scores, particularly for the 'no' class, suggesting some limitations in distinguishing non-smoke instances. While it exhibited high precision for the 'yes' class, its recall was relatively low for the 'no' class.
d. GBoost Excellence: GBoost performed exceptionally well with an accuracy of 1.00. Like other algorithms, it achieved perfect precision, recall, and F1 scores for both classes, signifying its reliability in smoke detection.

In summary, the findings indicate that CNN, SVM, Decision Tree, KNN, and GBoost can achieve high accuracy and precision in smoke detection. However, GNB showed limitations in distinguishing non-smoke instances, while the other algorithms exhibited robust performance across both classes. These results provide valuable insights into the comparative effectiveness of these algorithms for addressing the smoke detection task, with CNN and SVM emerging as solid contenders.

5.2 Interpretation of The Result

The interpretation of the research results reveals several key insights in smoke detection employing diverse machine learning algorithms. Firstly, the consistently high accuracy and precision achieved by CNN, SVM, Decision Tree, KNN, and GBoost underscore their effectiveness in accurately classifying instances with and without smoke, aligning with their well-established capabilities in classification tasks. Nevertheless, the underperformance of GNB, particularly in discerning non-smoke instances, raises questions. This unexpected outcome may be attributed to GNB's inherent feature independence assumption, which may not hold for the intricate visual features associated with smoke detection. This underscores the importance of algorithm selection aligned with data characteristics. These findings align with previous research, emphasizing the effectiveness of CNN, SVM, Decision Tree, KNN, and GBoost and GNB's sensitivity to feature independence assumptions. Alternative explanations, such as suboptimal hyperparameter tuning or incomplete data preprocessing, should be explored to understand the unexpected results. In conclusion, the study reaffirms the suitability of CNN, SVM, Decision Tree, KNN, and GBoost for smoke detection while highlighting GNB's limitations. These insights stress the significance of careful algorithm selection and parameter tuning in machine-learning tasks and offer valuable directions for future research in smoke detection and image classification.

5.3 Implication of the research

The implications of this research are far-reaching, given the critical relevance of smoke detection in various domains. By evaluating and comparing the performance of CNN with conventional machine learning algorithms for smoke detection, this study provides valuable insights into selecting appropriate methodologies for addressing this vital task. The results of this research align with previously discussed literature and existing knowledge, showcasing that CNN, SVM, Decision Tree, KNN, and GBoost can offer high precision and recall in smoke detection. The study highlights the importance of algorithm selection based on the specific characteristics of the data and use case. Additionally, it underlines the potential of CNN as a robust solution, offering new insights into applying deep learning techniques for smoke detection, with implications for enhancing safety measures, environmental monitoring, and regulatory compliance efforts.

5.4 Limitations of The Research

The limitations of this research should be considered to provide a comprehensive understanding of the study's scope. While the findings suggest the effectiveness of CNN and other machine learning algorithms in smoke detection, it's essential to acknowledge the limitations inherent in the dataset used and the specific experimental setup. The dataset, while substantial, may not fully represent the diversity of real-world scenarios for smoke detection. Additionally, variations in data quality or preprocessing techniques may impact the results. However, despite these limitations, the results remain valid for addressing the research questions, as the study offers valuable insights into algorithm selection and performance, which can guide the selection of appropriate methodologies for smoke detection tasks in various practical applications.

5.5 Future Research Recommendation

In light of this research, several recommendations for practical implementations and future research avenues emerge. Firstly, the findings affirm the suitability of CNN and other machine learning algorithms for smoke detection, suggesting their application in real-world scenarios like fire prevention, environmental monitoring, and safety enforcement. Future implementations could involve integrating these models into smart surveillance systems or IoT devices for real-time smoke detection. Furthermore, future research could delve into optimizing hyperparameters and exploring advanced neural network architectures to enhance the performance of CNN in this context. Additionally, investigating ensemble learning techniques and multi-sensor data fusion for more robust smoke detection systems presents exciting prospects. Moreover, research on developing transferable models for different environments and scenarios can contribute to broader applicability. Lastly, exploring the incorporation of temporal data or video frames for smoke detection in dynamic settings could be a valuable avenue for future investigation, ensuring the continued advancement of smoke detection technology.

6. Conclusion

The analysis of the classification algorithms demonstrates that CNN achieves perfect performance, with an accuracy of 1.00 and precision, recall, and F1-scores of 1.00 for both the "no" and "yes" classes. This indicates that CNN is highly effective and reliable for this classification task. Similarly, SVM, Decision Tree, KNN, and Gradient Boosting (Gboost) also achieve perfect performance across all metrics, matching CNN's results. However, Gaussian Naive Bayes (GNB) underperforms significantly, with an accuracy of 0.78 and lower precision, recall, and F1-scores, particularly for the "no" class (F1-score of 0.43).

CNN performs on par with other top-performing algorithms like SVM, Decision Tree, KNN, and Gboost, showcasing its robustness and suitability for this task. GNB, on the other hand, lags behind, highlighting its limitations in handling the dataset effectively. These results suggest that CNN is a strong choice for classification tasks requiring high accuracy and reliability, especially when compared to weaker performers like GNB. Future work could explore the scalability and computational efficiency of these algorithms to further refine their applicability in real-world scenarios.

References

[1] A. Setiono et al., “Performance of an Electrothermal MEMS Cantilever Resonator with Fano-Resonance Annoyance under Cigarette Smoke Exposure,” Sensors, vol. 21, no. 12, 2021, doi: 10.3390/s21124088.

[2] J. B. Sanger, L. Sitanayah, and V. D. Kumenap, “Detection System for Cigarette Smoke,” in 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Nov. 2019, pp. 145–149. doi: 10.1109/ICITISEE48480.2019.9003901.

[3] Somantri, R. F. Ridwanullah, Hendra, and D. Safitri, “Cigarette Smoke Detection System for Non-Smoking Areas Based on IoT and Face Recognition,” in 2020 6th International Conference on Computing Engineering and Design (ICCED), Oct. 2020, pp. 1–6. doi: 10.1109/ICCED51276.2020.9415798.

[4] X. Na, L. Zhang, H. Wang, and M. Tan, “Adverse effect assessment of fluorescent carbon dots in cigarette smoke,” NanoImpact, vol. 19, p. 100241, Jul. 2020, doi: 10.1016/j.impact.2020.100241.

[5] M. Jeong et al., “Public Understanding of Cigarette Smoke Chemicals: Longitudinal Study of US Adults and Adolescents,” Nicotine Tob. Res., vol. 22, no. 5, pp. 747–755, Apr. 2020, doi: 10.1093/ntr/ntz035.

[6] L.-Y. Xu, Y. Li, X.-X. Li, and J.-H. He, “Detection of cigarette smoke using a fiber membrane filmed with carbon nanoparticles and a fractal current law,” Therm. Sci., vol. 24, no. 4, pp. 2469–2474, 2020, doi: 10.2298/TSCI2004469X.

[7] K. J. Heo, G. D. Lee, S. J. Doh, and J. H. Jung, “Effect of cigarette smoke on the lifetime of electret air filters,” Sci. Total Environ., vol. 807, p. 150754, Feb. 2022, doi: 10.1016/j.scitotenv.2021.150754.

[8] M. C. B. Araújo and M. F. Costa, “From Plant to Waste: The Long and Diverse Impact Chain Caused by Tobacco Smoking,” Int. J. Environ. Res. Public. Health, vol. 16, no. 15, 2019, doi: 10.3390/ijerph16152690.

[9] R. Prieux, M. Eeman, B. Rothen-Rutishauser, and G. Valacchi, “Mimicking cigarette smoke exposure to assess cutaneous toxicity,” Toxicol. In Vitro, vol. 62, p. 104664, Feb. 2020, doi: 10.1016/j.tiv.2019.104664.

[10] Armando Peruga et al., "Compliance with the smoking ban in enclosed, semi-open and open areas of workplaces and public places in Chile," Tob. Control, vol. 30, no. 5, p. 570, Sep. 2021, doi: 10.1136/tobacco control-2020-055632.

[11] A. Chaudhary et al., “An assessment of compliance with the provisions of Cigarettes and Other Tobacco Products Act 2003: Is Shimla a smoke-free city?,” Natl. Med. J. India, vol. 33, no. 6, p. 335, 2020, doi: 10.4103/0970-258X.321141.

[12] M. Bodas and K. Peleg, “Pandemic Fatigue: The Effects Of The COVID-19 Crisis On Public Trust And Compliance With Regulations In Israel,” Health Aff. (Millwood), vol. 40, no. 8, pp. 1225–1233, Aug. 2021, doi: 10.1377/hlthaff.2021.00171.

[13] A. Karpilow, R. Cherkaoui, S. D’Arco, and T. D. Duong, “Detection of Bad PMU Data using Machine Learning Techniques,” in 2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Feb. 2020, pp. 1–5. doi: 10.1109/ISGT45199.2020.9087782.

[14] J. Waring, C. Lindvall, and R. Umeton, “Automated machine learning: Review of the state-of-the-art and opportunities for healthcare,” Artif. Intell. Med., vol. 104, p. 101822, Apr. 2020, doi: 10.1016/j.artmed.2020.101822.

[15] X. Cheng, “Research on Application of the Feature Transfer Method Based on Fast R-CNN in Smoke Image Recognition,” Adv. Multimed., vol. 2021, p. 6147860, Nov. 2021, doi: 10.1155/2021/6147860.

[16] D. Sheng, J. Deng, and J. Xiang, “Automatic Smoke Detection Based on SLIC-DBSCAN Enhanced Convolutional Neural Network,” IEEE Access, vol. 9, pp. 63933–63942, 2021, doi: 10.1109/ACCESS.2021.3075731.

[17] Z. Wang, L. Wu, T. Li, and P. Shi, “A Smoke Detection Model Based on Improved YOLOv5,” Mathematics, vol. 10, no. 7, 2022, doi: 10.3390/math10071190.

[18] M. M. Hasan, N. Hasan, M. S. A. Alsubaie, and M. M. R. Komol, “Diagnosis of Tobacco Addiction using Medical Signal: An EEG-based Time-Frequency Domain Analysis Using Machine Learning,” Adv. Sci. Technol. Eng. Syst. J., vol. 6, no. 1, pp. 842–849, Feb. 2021, doi: 10.25046/aj060193.

[19] M. U. Hashmi, N. D. Nguyen, M. Johnstone, K. Backholer, and A. Bhatti, “Application Based Cigarette Detection on Social Media Platforms Using Machine Learning Algorithms,” in Future Data and Security Engineering, T. K. Dang, J. Küng, T. M. Chung, and M. Takizawa, Eds., Cham: Springer International Publishing, 2021, pp. 68–80.
[20] J. Lin and H. Guo, “Application of Machine Learning in Cigarette Quality Evaluation Method,” J. Phys. Conf. Ser., vol. 1607, no. 1, p. 012011, Aug. 2020, doi: 10.1088/1742-6596/1607/1/012011.
[21] C. Shen, J. Hu, C. Wang, and D. Li, “LFNet: Lightweight fire smoke detection for uncertain surveillance environment,” in 2020 6th International Conference on Big Data and Information Analytics (BigDIA), Dec. 2020, pp. 11–17. doi: 10.1109/BigDIA51454.2020.00011.
[22] Li Liu, Shi-jie Xu, and Song-zhan Li, “Detection of reactive oxygen species in mainstream cigarette smoke by a fluorescent probe,” presented at the Proc.SPIE, Aug. 2009, p. 738236. doi: 10.1117/12.835678.
[23] C. Cui and R. Xu, “Multiple Machine Learning Algorithms for Human Smoking Behavior Detection,” in 2022 International Conference on Machine Learning and Intelligent Systems Engineering (MLISE), Aug. 2022, pp. 240–244. doi: 10.1109/MLISE57402.2022.00054.
[24] J. Liao, J. Zou, A. Shen, J. Liu, and X. Du, “Cigarette end detection based on EfficientDet,” J. Phys. Conf. Ser., vol. 1748, no. 6, p. 062015, Jan. 2021, doi: 10.1088/1742-6596/1748/6/062015.
[25] R. Kaabi, M. Bouchouicha, A. Mouelhi, M. Sayadi, and E. Moreau, “An Efficient Smoke Detection Algorithm Based on Deep Belief Network Classifier Using Energy and Intensity Features,” Electronics, vol. 9, no. 9, 2020, doi: 10.3390/electronics9091390.
[26] Y. Luo, L. Zhao, P. Liu, and D. Huang, “Fire smoke detection algorithm based on motion characteristic and convolutional neural networks,” Multimed. Tools Appl., vol. 77, no. 12, pp. 15075–15092, Jun. 2018, doi: 10.1007/s11042-017-5090-2.
[27] P. Teerarassamee, R. Chanklan, K. Kerdprasop, and N. Kerdprasop, “Smoke Detection with Ensemble Modeling,” Int. J. Model. Optim., pp. 44–47, May 2023, doi: 10.7763/IJMO.2023.V13.823.
[28] H. Lu et al., “Research on cigarette size detection based on machine vision,” in 2023 IEEE 3rd International Conference on Electronic Technology, Communication and Information (ICETCI), May 2023, pp. 1050–1053. doi: 10.1109/ICETCI57876.2023.10176835.
[29] Moyuan Li, Xiaobin Xu, Qinyang Qu, and Jiali Wang, “Research on object detection method in smoke and fog environment,” presented at the Proc.SPIE, Apr. 2023, p. 126176D. doi: 10.1117/12.2666704.
[30] S. Sun et al., “Research on Cigarette Shape Detection based on Dual Channel Acquisition,” Highlights Sci. Eng. Technol., vol. 15, pp. 5–9, Nov. 2022, doi: 10.54097/hset.v15i.2193.
[31] J. Xie, F. Yu, H. Wang, and H. Zheng, “Class Activation Map-Based Data Augmentation for Satellite Smoke Scene Detection,” IEEE Geosci. Remote Sens. Lett., vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2022.3179013.
[32] F. R. Fadhil and A. P. W. Wibowo, “SMOKE DETECTION ON CNN BASED VIDEO SURVEILLANCE SYSTEM,” J. Darma Agung, vol. 31, no. 1, p. 377, Mar. 2023, doi: 10.46930/ojsuda.v31i1.3010.
[33] Y. Zheng, G. Zhang, S. Tan, Z. Yang, D. Wen, and H. Xiao, “A forest fire smoke detection model combining CNN and vision transformer,” Front. For. Glob. Change, vol. 6, p. 1136969, Apr. 2023, doi: 10.3389/ffgc.2023.1136969.
[34] C. Bahhar et al., “Wildfire and Smoke Detection Using Staged YOLO Model and Ensemble CNN,” Electronics, vol. 12, no. 1, 2023, doi: 10.3390/electronics12010228.
[35] J. Zulkarnain, M. R. Pahlevi, Y. Astica, W. Pangestuti, and K. Kusrini, “Fire Detection based on Smoke Image using CNN,” in 2022 4th International Conference on Cybernetics and Intelligent System (ICORIS), Oct. 2022, pp. 1–5. doi: 10.1109/ICORIS56080.2022.10031555.
[36] J. S. Almeida, C. Huang, F. G. Nogueira, S. Bhatia, and V. H. C. de Albuquerque, “EdgeFireSmoke: A Novel Lightweight CNN Model for Real-Time Video Fire–Smoke Detection,” IEEE Trans. Ind. Inform., vol. 18, no. 11, pp. 7889–7898, Nov. 2022, doi: 10.1109/TII.2021.3138752.
[37] T. Singh, “Real Time Cigarette Detection using Deep Learning Models,” Int. J. Res. Appl. Sci. Eng. Technol., vol. 11, no. 5, pp. 1946–1948, May 2023, doi: 10.22214/ijraset.2023.51977.
[38] J. S. Tomiazzi, D. R. Pereira, M. A. Judai, P. A. Antunes, and A. P. A. Favareto, “Performance of machine-learning algorithms to pattern recognition and classification of hearing impairment in Brazilian farmers exposed to pesticide and/or cigarette smoke,” Environ. Sci. Pollut. Res., vol. 26, no. 7, pp. 6481–6491, Mar. 2019, doi: 10.1007/s11356-018-04106-w.

[39] V. Maslej-Krešňáková, M. Sarnovský, and J. Jacková, “Use of Data Augmentation Techniques in Detection of Antisocial Behavior Using Deep Learning Methods,” Future Internet, vol. 14, no. 9, 2022, doi: 10.3390/fi14090260.
[40] H. Zhang, Z. Xu, X. Han, and W. Sun, “Data Augmentation Using Bitplane Information Recombination Model,” IEEE Trans. Image Process., vol. 31, pp. 3713–3725, 2022, doi: 10.1109/TIP.2022.3175429.
[41] P. Belsare, V. Y. Senyurek, M. H. Imtiaz, S. Tiffany, and E. Sazonov, “Computation of Cigarette Smoke Exposure Metrics From Breathing,” IEEE Trans. Biomed. Eng., vol. 67, no. 8, pp. 2309–2316, Aug. 2020, doi: 10.1109/TBME.2019.2958843.
[42] P. M. J. Bos, L. G. Soeteman-Hernández, and R. Talhout, “Risk assessment of components in tobacco smoke and e-cigarette aerosols: a pragmatic choice of dose metrics,” Inhal. Toxicol., vol. 33, no. 3, pp. 81–95, Feb. 2021, doi: 10.1080/08958378.2021.1909678.
[43] J. Tyrrell, A. Ghosh, N. D. Manzo, S. H. Randell, and R. Tarran, “Evaluation of chronic cigarette smoke exposure in human bronchial epithelial cultures,” J. Appl. Toxicol., vol. 43, no. 6, pp. 862–873, Jun. 2023, doi: 10.1002/jat.4430.
[44] H.-J. Haussmann, “Use of Hazard Indices for a Theoretical Evaluation of Cigarette Smoke Composition,” Chem. Res. Toxicol., vol. 25, no. 4, pp. 794–810, Apr. 2012, doi: 10.1021/tx200536w.
[45] M. J. Oldham, H.-J. Haussmann, W. Gomm, L. T. Rimmer, M. J. Morton, and W. J. McKinney, “Discriminatory power of standard toxicity assays used to evaluate ingredients added to cigarettes,” Regul. Toxicol. Pharmacol., vol. 62, no. 1, pp. 49–61, Feb. 2012, doi: 10.1016/j.yrtph.2011.11.018.
[46] K. Yamada and H. Maruta, “A Study on Smoke Detection Method with Long Short-Term Memory Network,” in IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society, Oct. 2019, pp. 54–59. doi: 10.1109/IECON.2019.8926941.
[47] M. F. Al-Rawi and Y. A. G. Alyouzbaki, “Novel technique based on cascade classifiers for smoke image detection,” Int. Rev. Appl. Sci. Eng., vol. 12, no. 3, pp. 293–300, Jul. 2021, doi: 10.1556/1848.2021.00282.
[48] Department of Electrical Engineering, National Chung Hsing University, Taiwan, T.-C. Chien, C.-C. Lin, and C.-P. Fan, “Deep Learning Based Driver Smoking Behavior Detection for Driving Safety,” J. Image Graph., vol. 8, no. 1, pp. 15–20, 2020, doi: 10.18178/joig.8.1.15-20.
[49] J. S. Tomiazzi, M. A. Judai, G. A. Nai, D. R. Pereira, P. A. Antunes, and A. P. A. Favareto, “Evaluation of genotoxic effects in Brazilian agricultural workers exposed to pesticides and cigarette smoke using machine-learning algorithms,” Environ. Sci. Pollut. Res., vol. 25, no. 2, pp. 1259–1269, Jan. 2018, doi: 10.1007/s11356-017-0496-y.
[50] Z. Zhao, G. Cui, D. Li, and Q. Chen, “Video-based smoke detection by using motion, color, and texture features,” in Third International Symposium on Computer Engineering and Intelligent Communications (ISCEIC 2022), X. Ye and X. Ben, Eds., Xi’an, China: SPIE, Feb. 2023, p. 65. doi: 10.1117/12.2660953.
[51] M. Hashemzadeh, N. Farajzadeh, and M. Heydari, "Smoke detection in video using convolutional neural networks and efficient spatiotemporal features," Appl. Soft Comput., vol. 128, p. 109496, Oct. 2022, doi: 10.1016/j.asoc.2022.109496.
[52] N. Asiri, O. Bchir, M. M. B. Ismail, M. Zakariah, and Y. A. Alotaibi, “Image-based smoke detection using feature mapping and discrimination,” Soft Comput., vol. 25, no. 5, pp. 3665–3674, Mar. 2021, doi: 10.1007/s00500-020-05396-4.
[53] Z. Liu, X. Yang, Y. Liu, and Z. Qian, “Smoke-Detection Framework for High-Definition Video Using Fused Spatial- and Frequency-Domain Features,” IEEE Access, vol. 7, pp. 89687–89701, 2019, doi: 10.1109/ACCESS.2019.2926571.
[54] M. Shah, K. Gandhi, K. A. Patel, H. Kantawala, R. Patel, and A. Kothari, “Theoretical Evaluation of Ensemble Machine Learning Techniques,” in 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT), Jan. 2023, pp. 829–837. doi: 10.1109/ICSSIT55814.2023.10061139.
[55] R. M. Shukla and J. Cartlidge, “Challenges Faced by Industries and Their Potential Solutions in Deploying Machine Learning Applications,” in 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC), Jan. 2022, pp. 0119–0124. doi: 10.1109/CCWC54503.2022.9720900.
[56] A. Arul, R. S. Hari Prakash, R. Gokul Raja, V. Nandhalal, and N. Sathish Kumar, "Fire Detection System Using Machine Learning," J. Phys. Conf. Ser., vol. 1916, no. 1, p. 012209, May 2021, doi: 10.1088/1742-6596/1916/1/012209.
[57] Departement of Informatics, Institut Teknologi Nasional Bandung (ITENAS), Jl. PHH. Mustafa No. 23 Bandung, Indonesia 40124, Y. I. Nurhasanah, M. M. Barmawi, Departement of

Informatics, Institut Teknologi Nasional Bandung (ITENAS), Jl. PHH. Mustafa No. 23 Bandung, Indonesia 40124, R. Prakarsa, and Departement of Informatics, Institut Teknologi Nasional Bandung (ITENAS), Jl. PHH. Mustafa No. 23 Bandung, Indonesia 40124, “Implementation of Mean of Maximum on Cigarette Smoke Control in a Room,” Electroteh. Electron. Autom., vol. 70, no. 3, pp. 59–68, Sep. 2022, doi: 10.46904/eea.22.70.3.1108006.

[58] S. Chaturvedi, P. Khanna, and A. Ojha, “A survey on vision-based outdoor smoke detection techniques for environmental safety,” ISPRS J. Photogramm. Remote Sens., vol. 185, pp. 158–187, Mar. 2022, doi: 10.1016/j.isprsjprs.2022.01.013.