energies

MDPI

Article

Data Fusion and Ensemble Learning for Advanced Anomaly Detection Using Multi-Spectral RGB and Thermal Imaging of Small Wind Turbine Blades

Majid Memari, Mohammad Shekaramiz *, Mohammad A. S. Masoum and Abdennour C. Seibi

Machine Learning and Drone Laboratory, Engineering Department, Utah Valley University, Orem, UT 84058, USA; mmemari@uvu.edu (M.M.); m.masoum@ieee.org (M.A.S.M.); aseibi@uvu.edu (A.C.S.)

* Correspondence: mshekaramiz@uvu.edu

Abstract: This paper introduces an innovative approach to Wind Turbine Blade (WTB) inspection through the synergistic use of thermal and RGB imaging, coupled with advanced deep learning techniques. We curated a unique dataset of 1000 thermal images of healthy and faulty blades using a FLIR C5 Compact Thermal Camera, which is equipped with Multi-Spectral Dynamic Imaging technology for enhanced imaging. This paper focuses on evaluating 35 deep learning classifiers, with a standout ensemble model combining Vision Transformer (ViT) and DenseNet161, achieving a remarkable 100% accuracy on the dataset. This model demonstrates the exceptional potential of deep learning in thermal diagnostic applications, particularly in predictive maintenance within the renewable energy sector. Our findings underscore the synergistic combination of ViT's global feature analysis and DenseNet161's dense connectivity, highlighting the importance of controlled environments and sophisticated preprocessing for accurate thermal image capture. This research contributes significantly to the field by providing a comprehensive dataset and demonstrating the efficacy of several deep learning models in ensuring the operational efficiency and reliability of wind turbines.

Keywords: deep learning; RGB imaging; thermal imaging; data fusion; wind turbine blades; fault detection; image classification; ensemble learning; structural integrity; inspection

Citation: Memari, M.; Shekaramiz, M.; Masoum, M.A.S.; Seibi, A.C. Data Fusion and Ensemble Learning for Advanced Anomaly Detection Using Multi-Spectral RGB and Thermal Imaging of Small Wind Turbine Blades. *Energies* **2024**, *17*, 673. https://doi.org/10.3390/en17030673

Academic Editor: Frede Blaabjerg

Received: 13 January 2024
Revised: 29 January 2024
Accepted: 29 January 2024
Published: 31 January 2024

1. Introduction

The global shift towards sustainable energy solutions has prominently included wind power, with wind turbines becoming ubiquitous onshore and offshore. These turbines, especially their blades, are engineering marvels designed to efficiently convert wind's kinetic energy into electricity [1]. However, the blades are not impervious to environmental stresses, which can lead to damage and necessitate the development of advanced detection and repair techniques to ensure their ongoing functionality and safety [2]. The dynamics of wind turbine performance, influenced by environmental intricacies such as wind veer and the wind shadow effect, highlight the complex relationship between renewable energy technology and atmospheric conditions [3].

Ensuring the integrity of Wind Turbine Blades (WTBs) is vital for the efficiency of wind energy systems. Defects arising from environmental wear, material fatigue, or manufacturing issues can significantly impact turbine performance. While traditional inspection methods are foundational, they often lack the capability to detect subtle or internal damages, prompting a need for more advanced diagnostic tools [2]. A significant advancement in this field is the use of thermal imaging. This technique not only allows for the detection of heat patterns caused by various defects such as delamination, cracks, and dents [4], but can also offer more safety by enabling inspections from a further distance. This distance-based approach can reduce the risks associated with close physical inspections, particularly in challenging or hazardous environments.

Further enhancing the capability of defect detection is the fusion of thermal imaging with RGB imagery. This approach combines the high-resolution detail of RGB images with the temperature variance detection of thermal images, creating a more comprehensive view [5]. Such a fusion enables precise localization of thermal anomalies and offers a detailed assessment of the blade's condition. It allows for a multi-dimensional analysis of the turbine blades, revealing not just the location (*Where*) but also the nature (*What*) of the potential issues [5]. Additionally, the use of thermal imaging in tandem with RGB imaging helps overcome some inherent limitations when these modalities are used individually, such as the sensitivity of thermal imaging to environmental conditions, which can lead to ambiguous results [6]. RGB images provide additional visual cues that help contextualize thermal anomalies and distinguish between true defects and false positives from reflective surfaces or variable temperatures [6].

Recent advancements, such as the work of Zhu et al., introduced novel techniques such as the Regression Cropdata-processing method and an adaptive feature fusion module for RGB and infrared images. These innovations have significantly enhanced the accuracy of defect detection, differentiating actual defects such as coating failures from false positives such as dust or organic residues. Their methods demonstrate substantial improvements in detection accuracy and precision, with the adaptive feature fusion module increasing the precision of detecting actual defects to 99% [7]. Zhu et al. also tackled the challenges of defect detection in WTBs by proposing a multi-feature fusion residual network, augmented by transfer learning. This approach has been shown to significantly reduce the training time while ensuring accurate defect detection in WTBs [8]. Further contributions in the field include Kwo et al.'s development of a calibration method for active optical Lock-in Thermography, which has been effectively used to detect various defects in WTBs [9]. Sanati et al. explored both passive and active thermography techniques, enhancing the quality of thermal images through Step Phase and Amplitude Thermography [4]. Manohar et al. utilized Infrared Thermography, particularly Lock-in Thermography, for the localization and sizing of deep-lying defects in WTBs. They have developed a 3D depth estimation model that addresses the limitations of classical depth estimation methods [10].

The integration of these advanced technologies has been further improved by the simultaneous capture of thermal and RGB images, made possible by recent sensor technology advancements [4]. Modern data processing capabilities, powered by artificial intelligence and machine learning algorithms, have further enhanced the efficiency and reliability of these diagnostic tools [4]. However, the nuanced interpretation of fused thermal and RGB images still requires expert knowledge and standardized methodologies to adapt to different environmental and operational conditions [11–16].

Despite the significant progress in WTB inspection techniques, there remain notable challenges that current methodologies struggle to address. One of the primary limitations is the lack of adaptability to varied environmental conditions. This shortcoming often leads to a decrease in the accuracy and reliability of defect detection under different operational scenarios. Traditional methods, which primarily focus on either thermal imaging or RGB imagery, fail to provide a comprehensive assessment of defects. Addressing these issues, our research introduces an innovative methodology that combines the strengths of data fusion and ensemble learning techniques. We integrate thermal imaging with high-resolution RGB imagery to enhance the precision and depth of anomaly detection. This combination not only improves the accuracy of identifying defects but also provides a robust framework that can adeptly adjust to environmental variations. Our method, therefore, offers a significant improvement over existing inspection techniques by being more adaptable and comprehensive. The integration of these two imaging techniques allows us to create a more detailed view of the WTBs. Thermal imaging excels in identifying temperature variances indicative of potential defects, while RGB imagery provides high-resolution visual details that assist in contextualizing these thermal anomalies. By fusing these two data sources, we can accurately locate and characterize defects, distinguishing between genuine issues and false positives caused by external factors such as reflective surfaces or inconsistent

environmental temperatures. Our approach also incorporates advanced machine learning algorithms, which further refine the analysis and interpretation of the fused imagery. These algorithms are trained to recognize patterns and anomalies, increasing the efficiency and accuracy of the inspection process. By leveraging ensemble learning, our method aggregates the insights from multiple models, ensuring a more reliable and accurate identification of defects. Overall, this advancement in WTB inspection represents a significant step forward in maintaining the structural integrity and performance efficiency of wind turbines. By offering a more accurate, efficient, and comprehensive solution, our methodology reinforces the role of wind turbines in sustainable energy production, ensuring their continued reliability and effectiveness in harnessing wind power.

Below, we delve into the diverse applications of deep learning, highlighting its transformative role in infrastructure inspection and maintenance across various energy sectors.

1.1. Deep Learning in WTB Defect Detection

Deep learning, a branch of machine learning, has revolutionized data processing by mimicking the human brain's neural networks. Utilizing multi-layered artificial neural networks allows models to extract complex patterns from large, unstructured data sets, finding significant use in image processing and classification [17]. This methodology's ability to learn directly from data has led to its widespread adoption in various fields [18]. Moving from the general concept of deep learning to a specific use case, Convolutional Neural Networks (CNNs) emerge as a key player in image-related tasks. CNNs excel in automatically learning spatial features. This capability makes them highly effective for object detection, facial recognition, and image segmentation, demonstrating the practical applications of deep learning in intricate tasks [17]. Deep learning has revolutionized infrastructure inspection across various industries. Particularly in the energy sector, deep learning can significantly improve the safety of inspection methods by enabling analysis of remotely captured thermal images. For example, in wind turbine inspections, deep learning models analyze thermal images to detect blade defects by identifying thermal patterns indicative of different types of damage, thereby optimizing the precision and efficiency of inspections [19]. In the context of solar farms, drones capture images that are processed by deep learning models to identify issues such as dirt accumulation, panel breakages, or other anomalies that could reduce the efficiency [20]. Similarly, deep learning models analyze images or videos captured by drones or robotic devices to detect rust, cracks, or leaks in oil and gas pipelines [21]. Aerial imaging is used to monitor power lines by drones equipped with deep learning algorithms. They also be used for sagging lines, vegetation encroachments, or damaged insulators [22]. The primary advantage of employing deep learning in these applications is its ability to process large volumes of data quickly and consistently, providing valuable insights for timely maintenance and repairs [23]. As the energy industry continues to evolve, integrating deep learning into regular inspection routines promises significant improvements in the safety, efficiency, and longevity of critical infrastructure [24]. Utilizing multi-layered artificial neural networks allows models to extract complex patterns from large, unstructured data sets, finding significant use in image processing and classification [17]. This methodology's ability to learn directly from data has led to its widespread adoption in various fields [18].

In the realm of wind turbine inspections, deep learning has been particularly transformative. For example, Wu developed an efficient and accurate damage detector for WTB images [25]. Zhang et al. explored image recognition of WTB defects using attention-based mobilenetv1-yolov4 and transfer learning [26], while Peng et al. worked on motion blur removal for drone-based WTB images using synthetic datasets [27]. Zhang and Wen introduced SOD-YOLO, a small target defect detection algorithm for WTBs based on improved YOLOv5 [28], demonstrating the depth of innovation in this field.

Other significant contributions include the works of Denhof et al., who conducted automatic optical surface inspection of wind turbine rotor blades using CNN [29]. Qiu et al. proposed an automatic visual defects inspection of WTBs via a YOLO-based small object

detection approach [30]. Similarly, Shaheed and Aggarwal analyzed wind turbine surface defect detection from drones using U-Net architecture [31]. Carnero et al. discussed a portable motorized telescope system for WTBs damage detection [32], and Zhou et al. focused on wind turbine actual defects detection based on visible and infrared image fusion [7].

1.2. Research Imperative and Study Objectives

In the field of renewable energy, ensuring the structural soundness of WTBs is a critical task. These blades are perpetually exposed to a variety of environmental and operational stresses. Therefore, the ability to detect defects efficiently, accurately, and quickly is essential for maintaining their operational effectiveness, safety, and durability. Traditional inspection methods, which typically rely on either RGB or thermal imaging, have significant limitations. RGB imaging offers high-resolution visual details but is ineffective in identifying subsurface irregularities. On the other hand, thermal imaging is adept at detecting temperature variations that may signify defects, but it lacks the necessary visual clarity to pinpoint these issues accurately.

To overcome these limitations, our research introduces a novel approach that synergizes multi-spectral imaging. We combine the high-resolution detail of RGB images with the thermal anomaly detection capabilities of thermal images. This integrated process enhances both the precision and comprehensiveness of our defect detection methodology. The incorporation of data fusion techniques further strengthens our analysis, allowing for more nuanced interpretation of the complex data derived from these two imaging types.

Our research also stands out by implementing ensemble learning. This approach integrates multiple predictive models to refine the accuracy of anomaly detection, particularly crucial for the diverse and intricate nature of WTB defects. The diverse nature of defects often requires a multifaceted detection approach, which singular imaging techniques struggle to provide. In addition, we conduct a comprehensive exploration of various deep learning architectures, including CNNs, ResNets, DenseNets, Inceptions, and transformer-based models. These architectures are pivotal in effectively processing and interpreting the multimodal data from RGB and thermal images. Our selection criteria prioritize reducing the risk of misclassification of defects, which carry significant safety and economic implications.

A key consideration in our approach is operational efficiency. We aim to minimize inspection downtime of wind turbine and the resultant lack of energy production, which is particularly crucial for large wind farms. Our focus is on models that strike a balance between high diagnostic accuracy and computational efficiency. This is in alignment with the latest advancements in edge computing, promising significant cost savings and enhancing the practicality of regular turbine maintenance. The objectives and contributions of this research are as follows:

(a) Objectives of This Research:
    1. Conduct an in-depth evaluation and comparison of existing deep-learning classifiers for WTB inspection.
    2. Develop a novel thermal imaging dataset, thereby enhancing the understanding of blade conditions.
    3. Generate industry recommendations based on our comparative analysis of deep learning classifiers.
    4. Identify areas for further research, aiming to bridge knowledge gaps in deep learning classifiers for thermal imaging-based inspections.

(b) Contributions of This Research:
    1. Introducing a synergistic use of Thermal and RGB Imaging for enhanced WTB inspection.
    2. Creating a novel thermal imaging dataset at UVU's Machine Learning and Drone Lab, featuring high-resolution images.
    3. Developing an ensemble learning model, integrating Vision Transformer (ViT) and DenseNet161, to achieve unparalleled classification accuracy.

The remainder of this paper is structured as follows: Section 2 details the methodology, including the data collection process and the specific deep learning models employed. Section 3 presents our experiments, results, and comparative analyses. The final section summarizes our findings and discusses future research directions, emphasizing the potential advancements and practical applications of our research in wind turbine maintenance.

1.3. Overview of the Deep Learning Models Used for Comparison

In this section, we explore the deep learning models that we have employed for the classification of thermal images of WTBs. The advancement of deep learning in image recognition tasks has paved the way for innovative diagnostic methods, and in this study, we delve into this sophisticated realm to unlock its potential for enhancing wind turbine structural integrity assessment.

Traditional architectures such as AlexNet and VGG have laid the groundwork in the domain of deep learning. However, their deeper iterations can sometimes be computationally extensive for certain tasks. In contrast, ResNets, through their innovative skip connections, enable the creation of deeper networks without the associated challenges of traditional architectures. For scenarios where model efficiency and deployment speed are paramount, compact architectures such as SqueezeNet and MobileNets emerge as ideal choices. DenseNets, with their unique feature-reuse mechanism, potentially offer enhanced accuracy with a reduced parameter count. On the other hand, recent architectures such as EfficientNet and ViTs stand out for their cutting-edge performance, albeit at the cost of increased complexity and computational demands.

CNNs, specifically designed for processing grid-like data such as images, are highly efficient in capturing spatial hierarchies and patterns. Their convolutional layers are adept at extracting features from images, recognizing textures, shapes, and other visual elements, which is vital for identifying anomalies in wind turbine blades. Conversely, Artificial Neural Networks (ANNs), lacking spatial feature extraction capabilities, are less suitable for image classification tasks that require an understanding of spatial hierarchies. Recurrent Neural Networks (RNNs), ideal for sequential data, do not align well with the spatial data structure in images and are less efficient in processing the high dimensionality of raw image data. This comparative analysis solidifies our choice of CNNs for classifying thermal images of wind turbine blades, given their architectural advantages in extracting and processing spatial features.

1.3.1. AlexNet

The inception of AlexNet in 2012 by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton marked a transformative era in the field of deep learning [33]. This model, a trailblazer in the ImageNet Large Scale Visual Recognition Challenge, dramatically reshaped image classification tasks and underscored the potential of deep learning, as depicted in Figure 1. AlexNet's architecture comprises five convolutional layers, each responsible for extracting progressively complex features, and three fully connected layers that culminate these features into specific class identifications [33]. It was among the first to adopt Rectified Linear Unit (ReLU) activations for enhanced training efficiency and introduced dropout layers as an innovative method to prevent overfitting [34,35]. Designed for optimal performance on two Nvidia GTX 580 GPUs, AlexNet set the foundation for leveraging GPU computing power in deep learning. The impact of AlexNet is profound, continuing to resonate across a myriad of deep learning models that have emerged since its introduction [33].

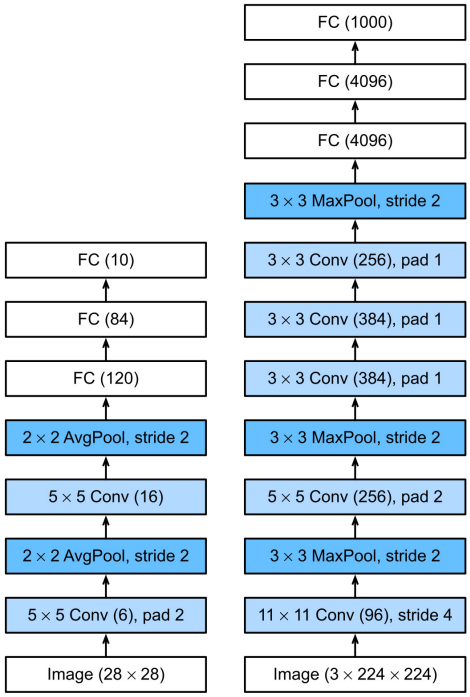


Figure 1. AlexNet architecture [36].

1.3.2. VGG

Developed by the Visual Geometry Group (VGG) at the University of Oxford in 2014, the VGG models have been instrumental in demonstrating the significance of depth in neural networks [37]. These models, featuring architectures with up to 19 layers, underscored the principle that a network's performance can be substantially enhanced by increasing its depth. A key characteristic of the VGG architecture is the adoption of 3×3 convolutional filters, which facilitate intricate feature detection while maintaining a manageable number of parameters [37]. The VGG family, encompassing variants such as VGG11, VGG13, VGG16, and VGG19, offers diversity in depth. This range allows for a balance between detailed feature representation and the risk of overfitting [37]. Some versions of the VGG models also integrate batch normalization, enhancing the efficiency and stability of the training process [38]. Despite their substantial computational requirements, VGG models are renowned for their profound depth and robust feature extraction capabilities, making them highly effective for transfer learning tasks [37].

1.3.3. Inception and Xception

The Inception architecture, also known as GoogLeNet, represents a significant advancement in neural network design. Developed by Google, its goal was to create a network that is deep and wide but maintains computational efficiency. A key feature of this architecture is the *Inception module*, as shown in Figure 2, which integrates concurrent pooling and convolutions of varying sizes (1×1, 3×3, and 5×5). This design allows the network to capture image features at multiple scales and abstraction levels. Mathematically, an Inception module can be represented as:

$$I(x) = [f_{1\times1}(x), f_{3\times3}(x), f_{5\times5}(x), f_{pool}(x)], \quad (1)$$

where $f_n(x)$ denotes a convolution operation with an $n \times n$ filter and $f_{pool}(x)$ is a pooling operation. These operations are performed in parallel and their outputs are concatenated.

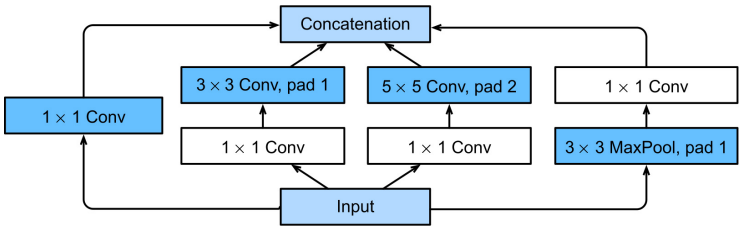


Figure 2. Inception block.

The subsequent versions, such as Inception v3 and v4, introduced factorized convolutions, asymmetric convolutions, grid-size reduction, and auxiliary classifiers to mitigate the vanishing gradient problem. Inception v3, for instance, emphasized computational efficiency by using 1×1 convolutions for dimensionality reduction and factorized larger convolutions for a balance between performance and efficiency.

Xception, termed *Extreme Inception*, is an evolution of the Inception model that sets itself apart by replacing the standard convolutions in the Inception modules with depthwise separable convolutions. This advanced convolution technique, essential to Xception, is designed to separately learn spatial features through depthwise convolution and channel-wise correlations through pointwise convolution. This approach not only enhances parameter efficiency but also improves performance in intricate image recognition tasks.

Depthwise separable convolution consists of two stages. The first stage, depthwise convolution, applies a single filter to each input channel. For an input feature map X of dimensions $H \times W \times D$ (height, width, depth), the depthwise convolution operation is represented as:

$$Y_{i,j,d} = \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} X_{i+m,j+n,d} \cdot F_{d,m,n} \,, \tag{2}$$

where $F_{d,m,n}$ denotes the filter for channel d. The second stage, pointwise convolution, then combines these features using 1×1 filters:

$$Z_{i,j,d'} = \sum_{d=0}^{D-1} Y_{i,j,d} \cdot F'_{d',d} \,, \tag{3}$$

producing the final output $Z_{i,j,d'}$, with $F'_{d',d}$ being the pointwise filter. The distinct phases of standard convolution, depthwise convolution, and pointwise convolution are visually demonstrated in Figure 3.

1.3.4. Residual Networks

The development of Residual Networks (ResNets) by Kaiming He and colleagues at Microsoft Research addressed the challenge of training deeper networks in deep learning [39]. ResNets introduced the concept of residual blocks, which focused on learning residual mappings to simplify the learning process [39]. The residual block allowed the network to adjust the identity mapping by a residual amount, achieved through skip connections that facilitated the flow of gradients and mitigated the vanishing gradient problem [39].

The core idea of a residual block is to learn the residual function $F(x)$ instead of the direct mapping $H(x)$. This residual function is defined as:

$$F(x) = H(x) - x, \tag{4}$$

where x is the input to the residual block, and $H(x)$ represents the desired output mapping of the network. The output of a residual block is the sum of the residual and the input, which can be expressed as:

$$\text{Output} = F(x) + x = \big(H(x) - x\big) + x = H(x) \tag{5}$$

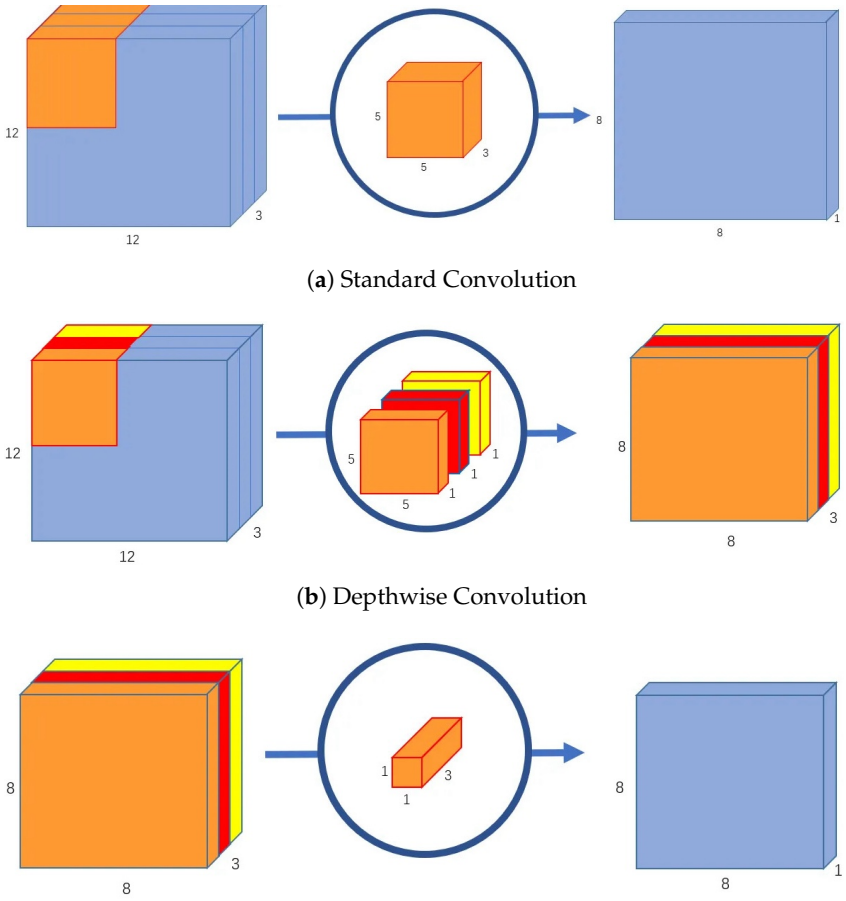


Figure 3. Illustration of the stages in depthwise separable convolution as used in Xception [40].

This formulation allows the network to learn identity mapping by default, facilitating the training of deeper models. Skip connections, also known as shortcuts, are a critical component of residual blocks, which enable the direct addition of the input of a layer to its output, as depicted in Figure 4. These connections allow for an unimpeded flow of gradients through the network, mitigating the vanishing gradient problem.

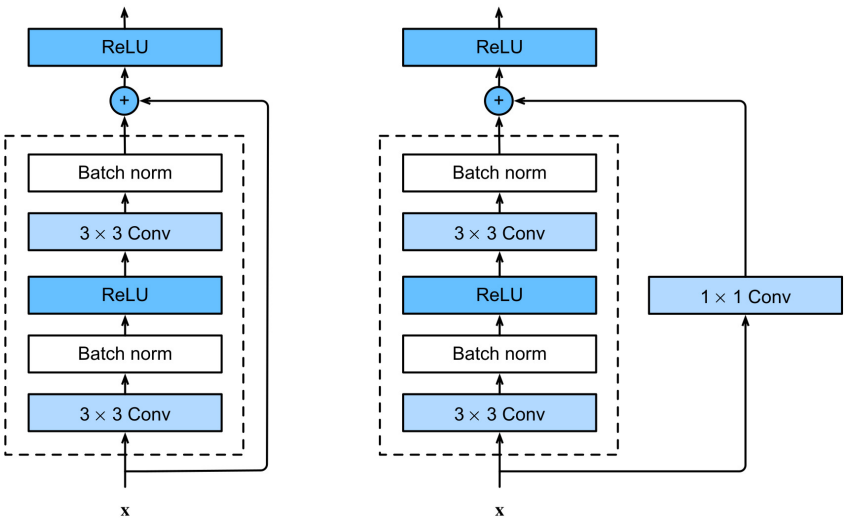


Figure 4. Residual blocks [36].

ResNets, or Residual Networks, offer a variety of architectural depths, with variants such as ResNet18, ResNet34, ResNet50, ResNet101, and ResNet152, where the numbers denote the count of layers in each model [39]. These deeper models, as depicted in Figure 5, are capable of discerning increasingly complex features, albeit at the cost of higher computational demands. Additionally, there exist *wide* versions of ResNets, which expand the number of filters in each layer, thereby enhancing the model's performance but also increasing its computational requirements [39].

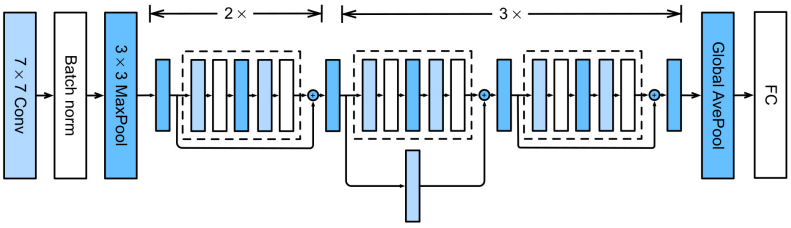


Figure 5. ResNet architecture [36].

The legacy of ResNets can be seen in their significant influence on the field of deep learning, inspiring subsequent models that manage both performance and computational resources, effectively [39].

1.3.5. Compact Architectures

SqueezeNet, a collaborative effort between DeepScale, the University of California, Berkeley, and Stanford University, aimed to reduce the number of parameters in the model while maintaining competitive accuracy. The key innovation of SqueezeNet is the Fire module, which addresses the high parameter demands of traditional convolutional layers. This module incorporates a squeeze convolutional layer to compress the input data, followed by an expanded layer that utilizes 1×1 and 3×3 filters to capture intricate patterns [41]. SqueezeNet has undergone significant evolution since its inception. The initial version, SqueezeNet 1.0, successfully reduced the model size to under 1 MB without compression [42]. SqueezeNet 1.1, its successor, introduced computational improvements that halved the required FLOPs without sacrificing accuracy. This lean architecture makes SqueezeNet well suited for deployment on edge devices with limited storage and computational capacity, offering expedited inference crucial for real-time applications [43].

1.3.6. Dense Connectivity Networks

DenseNets, short for Densely Connected Convolutional Networks, were introduced by Huang et al. [44]. Building on the concept of skip connections introduced by ResNets, DenseNets ensure that each layer receives inputs from all preceding layers, maximizing information flow within the network through dense inter-layer connectivity [44]. The distinctive feature of DenseNets is the *Dense Block* where each layer within the block ingests concatenated feature maps from all previous layers, in contrast to traditional models where a layer's input is solely the output of the preceding layer [44].

Dense connectivity in DenseNets offers several benefits. It improves the gradient flow during backpropagation, addressing the vanishing gradient problem in deep networks [44]. It also fosters feature reuse across layers, increasing network efficiency and reducing the number of required parameters [44,45]. This parameter-efficient model achieves high accuracy due to the comprehensive feature maps available from dense connections [46]. Additionally, DenseNets reduce redundancy, with layers learning new, complementary features on top of the accessible original features [47].

DenseNets come in various configurations denoted by the number following the model name, indicating the number of layers in the network, such as DenseNet121, DenseNet169, DenseNet201, and DenseNet161 [33,48]. To manage the feature map dimensions effectively as the network deepens, DenseNets incorporate 'transition layers' between dense blocks.

These transition layers, usually composed of batch normalization, a 1×1 convolution, and 2×2 average pooling, serve to control the feature map count and reduce their size, preparing them for the subsequent dense blocks. [49,50].

1.3.7. Mobile Architectures

The landscape of pervasive computing today, populated with an array of smart devices, IoT apparatuses, and edge computing nodes, necessitates machine learning models that are both powerful and lightweight. These models are designed to empower devices with limited computational resources with advanced deep learning functionalities. Among these, MobileNetV2 [51], MobileNetV3 [52], and EfficientNet [53] stand out as pioneering mobile architectures. Next, we discuss each version of MobileNet:

a. MobileNetV2 is the successor of MobileNet and is engineered to balance performance with computational efficiency. It introduces the concept of *Inverted Residuals*, where, unlike traditional residual blocks that increase and then decrease the channel dimensions, it starts with a slim depthwise convolution on the input and then uses a pointwise convolution to expand it. This technique helps in saving computational resources. Following the expansion, MobileNetV2 employs a linear bottleneck, maintaining the expanded channel size to prevent the loss of information that might occur due to non-linearities such as the ReLU function applied prematurely.
b. MobileNetV3 builds upon V2, integrating the principles of the latter with innovations that emerge from a combination of design finesse and automated architecture searches. It presents two versions: *Large* and *Small*, each tailored to different operational scenarios. The Large model prioritizes accuracy with a slight increase in computational demand, whereas the Small model focuses on conserving computational resources. This version of MobileNet advances with the incorporation of h-swish activation functions and leverages neural architecture search for enhanced efficiency optimization.
c. EfficientNet deviates from the traditional practice of manual architecture modifications or reliance on neural architecture search. Instead, it utilizes a systematic scaling approach that proportionally increases the network's width, depth, and input resolution, maintaining a balance in size and computation. EfficientNet's cornerstone is the compound coefficient, denoted as ϕ, which governs the uniform scaling across all dimensions. This coefficient is derived empirically to ensure the model grows harmoniously across its width, depth, and resolution. Starting with the base model EfficientNet-B0, increasing the compound coefficient leads to a series of variants, namely, B1, B2, and so on. Each increase aims to boost performance, but it also results in greater complexity and computational expense.

1.3.8. Vision Transformers (ViTs)

ViTs mark a paradigm shift in computer vision, challenging the dominance of CNNs [54]. Originating from the transformer architecture used in natural language processing, ViTs adapt these principles for image classification tasks [54,55]. The core of ViTs is the transformer architecture, originally designed for sequential data in natural language processing [54]. The attention mechanism is central to this architecture, allowing the model to focus on different segments of the input sequence. In the context of ViTs, this translates to focusing on various parts of an image [56]. The calculation of the attention mechanism in ViT is detailed in the formula below, and the attention mechanism used in ViT is illustrated in Figure 6. Additionally, the architecture of ViT is depicted in Figure 7.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{6}$$

where Q, K, and V represent the query, key, and value matrices, respectively. In this context, Q (query) refers to the matrix that contains the representation of the input that we are trying to find relevant information for. K (key) represents the matrix that we compare the

query against to find the relevant pieces of information. The term V (value) is the matrix containing the data that we actually want to retrieve. The dimension of the key, denoted as d_k, influences the scaling factor in the softmax calculation.

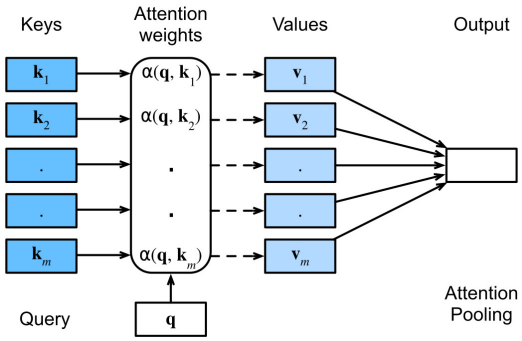


Figure 6. Attention Mechanism [36].

In ViTs, an image is divided into patches, akin to how text sequences are treated in NLP [57]. These patches are embedded into a sequence of tokens:

$$\text{Patch Embedding: } E = [E_1, E_2, \ldots, E_N], \tag{7}$$

where E_i represents the embedded vector for the i-th patch and N is the number of patches. These embedded patches are then processed through multiple layers of the transformer, capturing both local and global image contexts [58].

In our implementation, the *ViTForImageClassification* model from the *transformers* library is utilized [59]. This model processes images in patches (e.g., 16×16 pixels) with an input resolution of 224×224 pixels and is pre-trained on the ImageNet dataset [60]. The classifier layer of the model is adapted for a binary classification task, reducing the output neurons to two [61]. This demonstrates the adaptability of ViTs across different image classification scenarios.

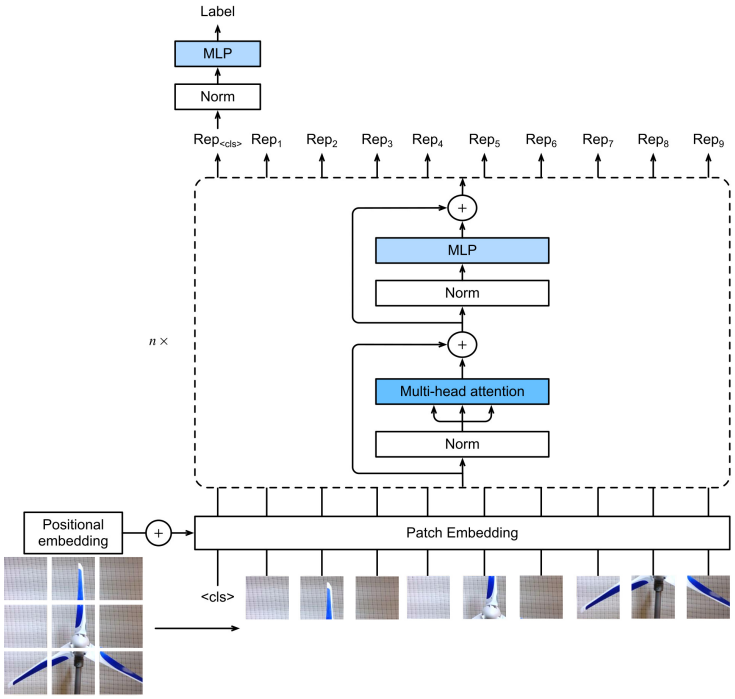


Figure 7. ViT architecture [36].

In the following section, we will delve into the specific experimental setups, including dataset details and the training process. We will also discuss the selection and configuration of models, the application of transfer learning, evaluation metrics, and a thorough analysis of the results. This will provide a comprehensive understanding of how our methods translate into tangible outcomes in the realm of WTB inspection.

2. Methodology

This section delves into the comprehensive methodology underpinning our research, focusing on the data collection process and the deployment of deep learning models for the classification of thermal images of WTBs. Our objective is to assess the capability of deep learning techniques in identifying issues related to the structural integrity of WTBs, an essential factor in predictive maintenance within the renewable energy sector.

2.1. MSX Technology

At the Machine Learning and Drone Lab of Utah Valley University, our research team is dedicated to enhancing a thermal imaging dataset for wind turbine blade (WTB) inspection. A key component in this endeavor is the FLIR C5 Compact Thermal Camera, as shown in Figure 8. This camera is especially notable for its Multi-Spectral Dynamic Imaging (MSX) technology, which is referenced in [62,63]. MSX technology plays a crucial role in combining thermal and standard RGB (Red, Green, Blue) imaging, thereby providing a more nuanced and comprehensive method for data collection and analysis. This paper delves into the intricacies of MSX technology, explores the dataset generated through this technology, and examines the models developed for differentiating between faulty and healthy images of WTBs.

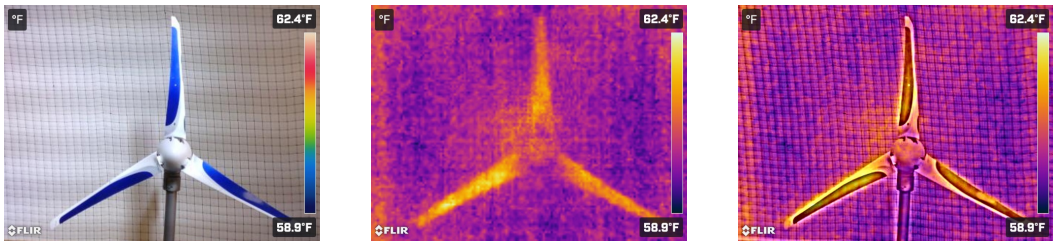
The MSX technology plays a pivotal role in integrating thermal and RGB imaging, enabling a more detailed and effective method of data acquisition and analysis. In this section, we delve into the intricacies of MSX technology, explore the dataset generated through this technology, and examine the models employed for categorizing images of WTBs as either faulty or healthy.

Figure 8. FLIR C5 camera, equipped with MSX technology.

MSX technology significantly enhances thermal imaging by superimposing high-resolution RGB imagery onto thermal data. This advanced technique combines detailed RGB visuals with the thermal data, yielding a composite image that is both intricate and informative. The integration of clear RGB imagery into the thermal images not only enriches the contextual understanding but also ensures the accuracy and integrity of the data sets.

The FLIR C5 camera was chosen not only for its MSX feature but also for its cost-effectiveness, making it an ideal choice within the constraints of our research. The MSX technology in the FLIR C5 facilitates the real-time integration of visual details with thermal images. This cost-effective solution is crucial for the accuracy of our dataset, providing genuine and consistent data without the need for post-capture fusion. The images produced are particularly useful for inspecting WTBs, as they comprehensively capture both thermal anomalies and essential visual details, crucial for an accurate assessment.

Thermal imaging technology captures infrared radiation emitted by objects, which is then translated into an image representing the temperature distribution of the area inspected. This technology is excellent for identifying heat-related issues but traditionally lacks the visual context of non-thermal features, making it challenging to interpret details in the environment [64–67]. Standard thermal imaging presents images based solely on temperature differences. While effective for detecting areas of heat and cold, it does not provide visual details, such as texture or specific shapes, which are often necessary for a complete understanding of the scene or object being inspected. MSX technology, utilized in devices such as the FLIR C5 Compact Thermal Camera, enhances standard thermal imaging by superimposing high-resolution RGB imagery onto thermal data [62]. This technology is designed to merge the visual clarity of RGB imaging with the functional insights of thermal imaging. MSX technology addresses the limitations of standard thermal imaging by embedding high-contrast, high-resolution details from an onboard digital camera onto thermal images in real-time. This integration allows for a more intuitive understanding of the thermal data, as the thermal image is overlaid with clear visual details [68,69]. The key advantage of MSX over standard thermal imaging is the enhanced feature recognition and clarity it provides. MSX images are not only thermally informative but also visually detailed, allowing for easier identification of objects, structural features, and potential issues in the thermal landscape. In practical applications, such as the inspection of WTBs, MSX technology proves to be superior [70,71]. The composite images generated by MSX offer a more comprehensive understanding of the condition of the blades, combining the thermal anomalies detected with visual cues that aid in pinpointing specific issues. Our dataset comprises 1000 thermal images, each with a resolution of 320 × 320 pixels, significantly benefiting from the MSX feature. These images are directly integrated with visual and thermal details during capture, thus preserving data authenticity. We have categorized the images into two sets: 500 images of healthy blades, and 500 faulty images with the forms of damage cracks, holes, and erosion. In our dataset, the incorporation of MSX and RGB–Thermal Fusion provides a composite view that enriches the images with contextual visual cues. These cues are vital for precise detection and categorization of anomalies via machine learning algorithms, enabling the training of robust models for identifying potential damages in WTBs. Figure 9 showcases examples of these visualizations.



(**a**) (**b**) (**c**)

Figure 9. Visualizations of WTBs in our experiments. (**a**) Digital RGB visualization of WTBs; (**b**) Thermal-only visualization of WTBs; (**c**) MSX visualization of WTBs.

Conscientious curation of the dataset included adjusting the background temperature settings in the FLIR C5 camera to compensate for the effect of external thermal sources. This, along with the camera's integrated MSX technology, ensured the accurate representation of the visual aspects of turbine blade conditions in the corresponding thermal images. External thermal sources that could affect WTBs include solar radiation, reflections from nearby structures, and ambient temperature variations. Controlled environmental settings were utilized to normalize these conditions and further minimize the noise in thermal data.

2.1.1. Handling Parallax Effects in MSX Imaging

The integration of MSX technology in the FLIR C5 camera inherently addresses several challenges associated with multi-spectral imaging, including the parallax effect. Parallax, a phenomenon where the position or direction of an object appears differently when viewed from different perspectives, can significantly impact the accuracy of merged thermal and RGB images. In the context of MSX technology, parallax could potentially lead to misalignments between thermal and RGB layers of the image. Such misalignments might obscure or falsely represent thermal anomalies, crucial for identifying defects in wind turbine blades. The FLIR C5's MSX technology is designed to minimize these parallax effects. It achieves this through real-time processing and intelligent alignment algorithms that overlay thermal data onto high-resolution RGB imagery. This process ensures the accurate representation of thermal anomalies in the context of the visual details provided by the RGB layer. Addressing parallax in MSX imaging is crucial for accurate defect detection. Misalignment caused by parallax could lead to incorrect interpretations of thermal anomalies, resulting in either missed defects or false positives. The MSX technology significantly contributes to minimizing these effects, thereby crucially improving the reliability and precision of our defect detection process. Our examination of the images generated by MSX illustrates the technology's effectiveness in addressing challenges related to parallax. The high level of accuracy in defect detection and localization across our dataset indicates the robustness of the MSX technology's parallax correction mechanism.

Each image in the dataset originates from the FLIR C5 camera's native 160 $\times$ 120 sensor resolution and is upscaled to 320 $\times$ 320 pixels using the FLIR Tools software, safeguarding the data fusion integrity. The camera's cloud connectivity features significantly expedited data management throughout the collection process. Utilizing RGB–Thermal Fusion, our dataset combines the strengths of both imaging modalities, enhancing object detection and situational awareness, especially in low visibility conditions and for edge detection. This fusion approach is proven to outperform traditional RGB models, especially in periods of limited visibility, making it ideal for rapid and accurate WTB integrity assessments [16]. Our approach ensures that the dataset is not only analytically robust but also immediately applicable for practical field inspections.

Figure 10 illustrates a sample set of RGB and thermal images from the wind turbine blades we studied. In these images, the blades are shown in a state of disrepair. More specifically, Figure 10a,b display sample images of a WTB with cracks. Figure 10a presents the blade in MSX format, highlighting the detailed visualization of the cracks, while Figure 10b shows the same blade in RGB format. The cracks are noticeably clear in the thermal image, with the thermal contrast sharply outlining the fissures. This highlights potential weak spots in the blade's structural integrity. Figure 10c,d focus on different damage types. Figure 10d illustrates the blade with erosion and holes in RGB format, whereas Figure 10c depicts these damages in Thermal Fusion format. The effectiveness of the Thermal Fusion image in revealing areas of material erosion or holes can be attributed to the changes in emissivity at these points. For example, cracks and holes often act as high-emissivity spots, resulting in a different thermal reading due to the contrast in emissivity between the damaged areas and the original material. This distinction is crucial for maintenance decisions, as the thermal imaging technology's sensitivity to emissivity variations facilitates the early detection of such defects and enables effective preventive maintenance strategies.

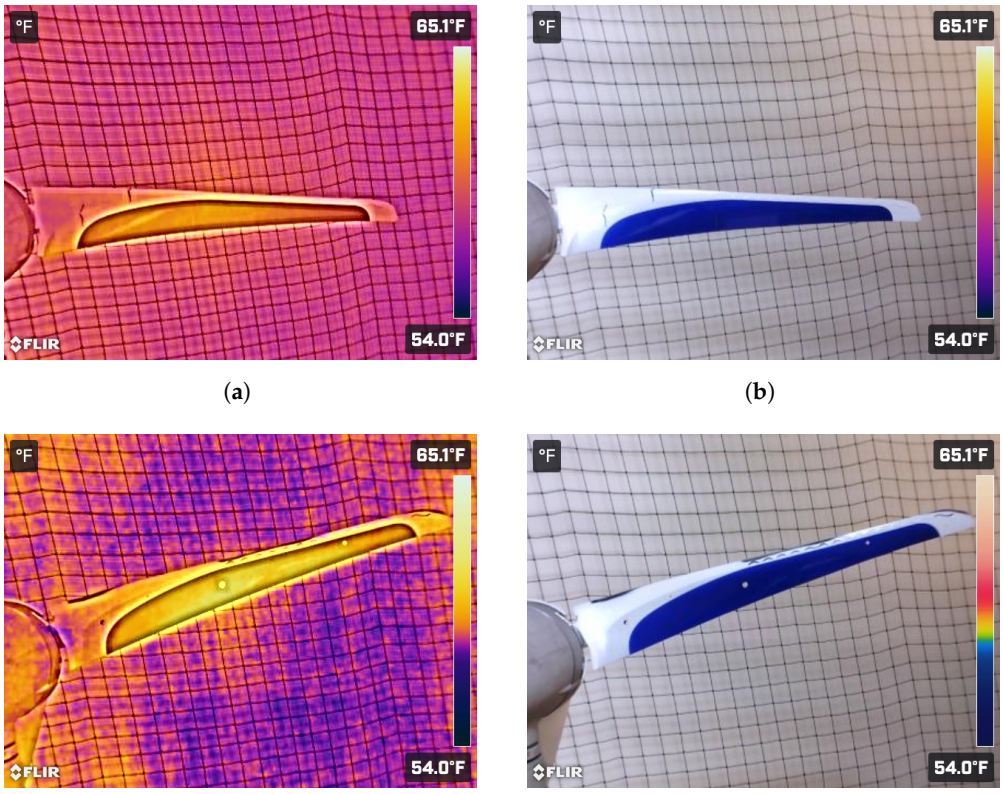


Figure 10. Visualization of WTB defects using multiple imaging techniques. (**a**) Sample MSX imaging of cracks; (**b**) Sample RGB imaging of cracks; (**c**) Sample MSX imaging of erosions and holes; (**d**) Sample RGB imaging of erosions and holes.

2.1.2. Dataset Preparation

The dataset underwent a rigorous refinement process. From the initial comprehensive dataset, we methodically divided it into training and validation subsets, ensuring both sets had a balanced representation of the classes. This 80–20% partitioning was instrumental in facilitating effective model training. In addition to these steps, our preprocessing approach involved several advanced techniques to enhance the robustness of our model, particularly for defect detection in thermal images captured by drones in the future.

To account for the varying orientations of defects that may occur in drone-captured imagery, we applied extensive rotation and flipping transformations to our dataset. This involved rotating each image at angles of 90°, 180°, and 270°, along with horizontal and vertical flipping. As a result, each original image was transformed into 16 unique variations, effectively increasing the diversity of our dataset by 1600%. This extensive augmentation ensures that our model is robustly trained to recognize defects in any orientation, which is critical for effective aerial surveillance applications. Recognizing the susceptibility of thermal imaging systems to real-world imperfections, we incorporated the injection of random Gaussian noise into the dataset. This augmentation was designed to make the model more resilient to the slight variations and imperfections typically encountered in field conditions, thereby improving its reliability in practical scenarios. Furthermore, given the propensity for motion blur occurring in images captured from drones, we introduced Gaussian blur as part of our preprocessing pipeline. This step is particularly important for mimicking the impact of drone movement and focus variations on the image quality. By training our model on these slightly blurred images, we aimed to reduce its sensitivity to sharpness and focus inconsistencies, which are common challenges in aerial thermal imaging. These preprocessing measures, while not explicitly detailed in our code, are indicative of the meticulous and comprehensive nature of our approach. They complement the initial steps of cropping, denoising, normalization, and resolution standardization,

ensuring that our models are provided with data of the highest quality and diversity, thus paving the way for effective training and robust performance.

3. Experiments and Results

In this section, we detail the comprehensive experiments conducted and the insightful results obtained in our endeavor to harness deep learning for the classification of thermal images of WTBs. Our experimental framework is designed to rigorously test various advanced neural network models, carefully evaluating their efficacy in accurately identifying structural anomalies in the blades. These experiments are not only crucial for validating our theoretical model configurations but also pivotal in demonstrating the practical applicability and effectiveness of our approach in real-world scenarios within the renewable energy sector.

3.1. Data Acquisition and Preprocessing Techniques

We curated a comprehensive dataset consisting of 1000 thermal images, each of which was rigorously processed to ensure uniformity and consistency across the dataset. This process involved several critical steps, including cropping, centering, and resizing each image to a uniform resolution of 320×320 pixels, as well as categorizing them as *healthy* or *faulty* for the purpose of WTB inspection. Our dataset is composed of 500 images of healthy blades and 500 images of faulty blades, ensuring a balanced representation between the two classes. The faulty images include various types of defects such as cracks, erosion, and holes, providing a comprehensive coverage of common fault types in wind turbine blades. During the preprocessing phase, we specifically focused on techniques suitable for thermal imaging, which included normalization to maintain consistent pixel intensity values across all images, and various augmentation techniques such as random rotation, flipping, noise injection, and blurring, designed to simulate the challenging conditions of real-world aerial photography.

3.2. Strategic Data Partitioning

In aligning with best practices in machine learning, we strategically partitioned our dataset into subsets designated for training, validation, and testing, with a distribution of 80%, 10% (of the training dataset), and 20%, respectively. This approach was specifically chosen to minimize the risk of overfitting and to validate the capability of our models to generalize effectively to new, unseen data.

3.3. Optimization of Model Parameters and Hyperparameter Tuning

Our methodology for refining the model's hyperparameters was comprehensive, ensuring the robustness and reliability of our ensemble model. We employed a two-pronged approach: initially applying a random search to explore a wide range of values, followed by Bayesian optimization to fine-tune and converge upon the most effective parameters. This dual approach allowed us to benefit from both the extensive search space of the random search and the precision of Bayesian optimization.

Specifically, the Bayesian optimization process was pivotal in determining the optimal threshold value for decision making in our ensemble model. We utilized a probabilistic model that was iteratively updated based on the performance metrics of previously evaluated hyperparameter sets. This method efficiently navigated the search space to hone in on the values that balanced the precision–recall trade-off most effectively. As a result, we found our decision threshold to be best set at 0.7, after testing values between 0.5 to 0.8, enhancing the model's predictive accuracy while maintaining an excellent balance between sensitivity and specificity.

The ranges for other key hyperparameters were selected as follows:

- **Learning Rate**: set between 1×10^{-5} and 1×10^{-4}, to fine-tune the speed of convergence against the stability of the training.

- **Batch Size**: varied between 16 and 128, to optimize computational resources and gradient estimation.
- **Optimizer**: chose the Adaptive Moment Estimation (Adam) for its efficiency in adjusting learning rates across parameters.
- **Weight Decay**: established within 1×10^{-4} to 1×10^{-2}, balancing adaptability and regularization to avert overfitting.
- **Early Stopping**: implemented to monitor validation loss and cease training when no improvement occurred over a set number of epochs.

This rigorous hyperparameter tuning process forms the foundation for the outstanding performance of the ensemble model in anomaly detection tasks.

3.4. Deep Learning Model Configurations

Our research involved a comprehensive range of deep learning models, each fine-tuned to analyze thermal images of WTBs for the purpose of identifying whether the blades were *healthy* or *faulty*. These models were methodically grouped based on the nature of their terminal layer, such as classifiers or fully connected layers, with notable examples like AlexNet, various iterations of VGG and DenseNet, and MobileNet. This grouping proved crucial for making precise fine-tunes for binary classification. Our ensemble included a spectrum of architectural styles, from classic convolutional networks like AlexNet and VGG to more sophisticated designs like ResNet and DenseNet, as well as cutting-edge models like ViT and EfficientNet. This varied selection enabled an in-depth examination of different image classification strategies and their effectiveness in assessing the condition of WTBs. A key element of our methodology was tailoring each model for binary classification to distinguish unequivocally between *healthy* and *faulty* blades. We employed a structured approach in setting up the models, using a *base_config* dictionary as a central tool to systematically organize the models based on their architecture and the adaptations needed for the end layer modifications.

3.5. Transfer Learning Approach

In our research focused on detecting anomalies on wind turbine blades, we employed a transfer learning strategy that utilizes the extensive pre-trained knowledge from models based on the ImageNet dataset. ImageNet, known for its vastness and diversity, has been a cornerstone in the advancement and benchmarking of cutting-edge deep learning models for image classification. By pre-training our models on ImageNet, we laid a solid foundation for them to recognize and understand various image features. This foundational knowledge is especially advantageous in handling the complexities and specific characteristics of thermal images used in wind turbine inspections [60,72].

Reduced Training Time with ViT

The application of ViT model in our study significantly reduced the training time. This efficiency is primarily due to ViT's prior training on the extensive and diverse ImageNet dataset, enabling it to start with a robust understanding of various image features. As a result, the adaptation of ViT to our specific task of anomaly detection in wind turbine blades required less time compared to training a model from scratch. This reduction in training time is a key benefit of transfer learning, particularly when dealing with large and complex datasets as in our case. Furthermore, ViT's ability to process images in patches and its attention mechanism, which focuses on different segments of the input, contributed to its rapid and effective adaptation to our task.

To further elaborate on this approach, the subsequent points outline the key stages and methods we implemented in our study. These include the initial phase of pre-training on the ImageNet dataset, followed by detailed fine-tuning and layer adjustments, all of which are integral to improving our ability to detect anomalies in wind turbine blades.

a. Pre-Training on ImageNet: Each model used in our study was initially pre-trained on the ImageNet dataset. This pre-training involved exposing the models to a large

variety of images, allowing them to learn a wide range of features, from basic shapes and textures to more complex patterns. This process endowed the models with a significant degree of generalized image recognition capability, which forms an essential basis for further specialization in our specific task.

b. Transfer Learning Implementation: We employed the transfer learning technique by fine-tuning these pre-trained models on our dataset of thermal images of WTBs. This approach allowed us to emphasize the learned features from ImageNet while adapting the models to the nuances of our specific application. The fine-tuning involved partial retraining of the models, where the initial layers, responsible for capturing basic image features, were kept frozen, and the deeper layers were trained to align with our task-specific features.

c. Layer Adjustment Strategies: A crucial aspect of our transfer learning approach was the adjustment of the final layers of the models to suit our binary classification task. This entailed transforming the output layers to yield predictions in a binary format, corresponding to the *healthy* and *faulty* conditions of the turbine blades. Additionally, we experimented with varying the number of layers to be retrained and the extent of their training, striking a balance between retaining learned features and adapting to new data.

Through this transfer learning approach, we effectively utilized the vast knowledge base of ImageNet to enhance the performance of our models on a specialized task. This methodology not only accelerated the training process, but also improved the models' accuracy and generalization capabilities in distinguishing between healthy and faulty turbine blades.

3.6. Results Analysis

This subsection delves into the detailed results obtained from the application of various deep learning models on our dataset, with a focus on their performance across multiple metrics. Table 1 shows the performance metrics of 35 implemented models, highlighting their accuracy, precision, recall, and F1-score. The F1-score is a harmonic mean of precision and recall, providing a single metric that balances both the false positives and false negatives. It is particularly useful in situations where an uneven class distribution might render metrics like accuracy. The F1-score ranges from 0 to 1, with a higher value indicating better model performance and a more balanced trade-off between precision and recall. Table 1 orders the models by their performance, demonstrating the comparative effectiveness of each architecture in the context of image classification of *healthy* and *faulty* for WTBs. Particularly notable are the performances of the ViT and DenseNet models, which stand out with high accuracy and F1-scores, exemplifying their precision and effectiveness in this task.

The F1-score is calculated as follows:

$$\text{F1-score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{8}$$

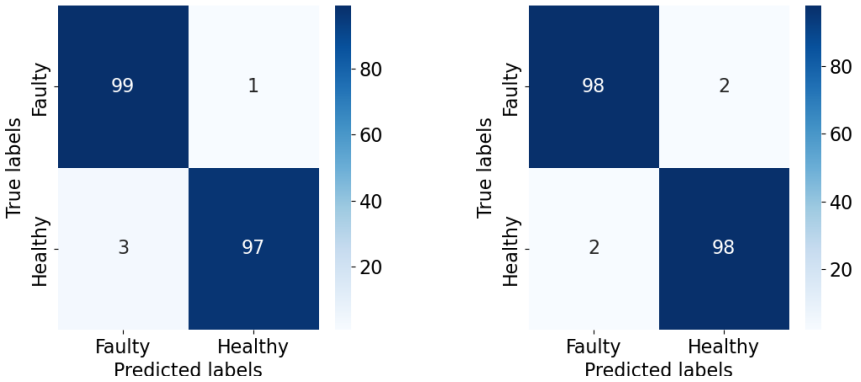
Table 1. Average performance metrics of implemented models after 10 random training runs (in %).

Performance	Model Name	Avg. Accuracy (%)	Avg. Precision (%)	Avg. Recall (%)	Avg. F1-Score (%)
1	**The Ensemble Model**	**100**	**100**	**100**	**100**
2	**ViT**	**97.98**	**97.97**	**97.99**	**97.98**
3	**densenet161**	**97.95**	**98.93**	**97.03**	**97.96**
4	resnet152	97.87	97.89	96.95	97.42
5	densenet201	97.45	96.97	97.94	97.46
6	wide_resnet50_2	97.48	96.13	98.96	97.52
7	vgg16	97.49	97.02	97.99	97.50
8	densenet169	97.02	95.21	99.02	97.08

Table 1. *Cont.*

Performance	Model Name	Avg. Accuracy (%)	Avg. Precision (%)	Avg. Recall (%)	Avg. F1-Score (%)
9	vgg13	97.03	96.10	98.02	97.05
10	vgg19_bn	96.02	97.94	94.02	95.94
11	vgg11	96.03	95.12	97.02	96.06
12	wide_resnet101_2	96.04	94.25	98.03	96.10
13	densenet121	96.03	95.12	97.02	96.06
14	resnet101	96.52	96.06	97.02	96.54
15	vgg11_bn	96.52	95.17	98.02	96.57
16	vgg16_bn	96.52	95.17	98.02	96.57
17	efficientnet_b1	94.82	93.72	96.02	94.87
18	mobilenet_v2_0.75	94.77	91.52	98.52	94.87
19	inception_v4	95.27	93.82	96.72	95.25
20	squeezenet1_0	95.52	95.98	95.02	95.50
21	resnet18	95.52	98.94	92.02	95.36
22	vgg13_bn	95.52	91.76	95.71	95.71
23	efficientnet_b0	94.52	94.08	95.02	94.55
24	vgg19	94.52	96.86	92.02	94.38
25	resnet50	94.52	93.22	96.02	94.59
26	mobilenet_v2_0.5	94.02	90.52	98.02	94.12
27	mobilenet_v3_medium	93.52	90.02	97.52	93.62
28	xception	95.02	93.29	97.02	95.12
29	resnet34	95.02	91.69	99.02	95.21
30	mobilenet_v2	95.02	91.69	99.02	95.21
31	inception_v3	95.02	96.89	93.02	94.92
32	mobilenet_v3_large	92.52	89.01	97.02	92.84
33	squeezenet1_1	91.02	88.70	94.02	91.28
34	mobilenet_v3_small	89.02	83.07	98.02	89.93
35	alexnet	88.98	83.61	97.02	89.79

Figure 11a,b illustrate the confusion matrices for the DenseNet and ViT models, respectively. These matrices provide a detailed breakdown of the classifiers' performance, showing the true positives, true negatives, false positives, and false negatives. The high accuracy of these models is reflected in the significant number of correct predictions.



(**a**) Confusion Matrix for DenseNet (**b**) Confusion Matrix for ViT

Figure 11. Confusion matrices of the top two models.

The Receiver Operating Characteristic (ROC) curves for DenseNet and ViT are depicted in Figures 12 and 13, respectively. ROC curves are essential tools in evaluating the performance of classification models, as they provide a visual representation of the trade-off between the True Positive Rate (TPR) and the False Positive Rate (FPR) across various threshold settings. This trade-off is crucial in determining the balance between correctly identifying positive cases and avoiding false alarms. For each model, the ROC curve plots the TPR

against the FPR at different classification thresholds, providing insight into the classifier's performance under varying conditions. The Area Under the Curve (AUC) of these ROC plots serves as a quantifiable measure of the model's discriminative ability. A higher AUC value indicates a better performance of the model in distinguishing between the two classes—in this case, the *healthy* and *faulty* conditions of the wind turbine blades. The AUC is a valuable metric as it summarizes the ROC curve into a single number, enabling a straightforward comparison between different models. A model with an AUC close to 1 demonstrates excellent classification ability with a high true positive rate and a low false positive rate, whereas a model with an AUC closer to 0.5 suggests performance no better than random chance.

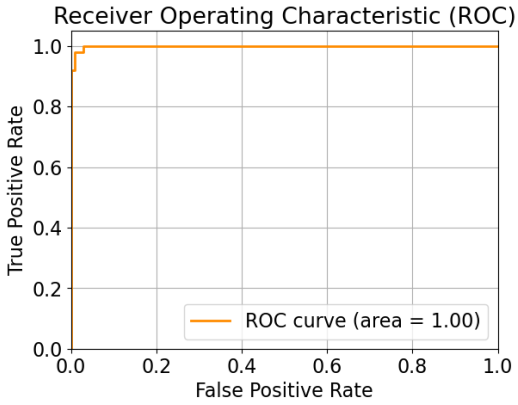


Figure 12. ROC curve for DenseNet.

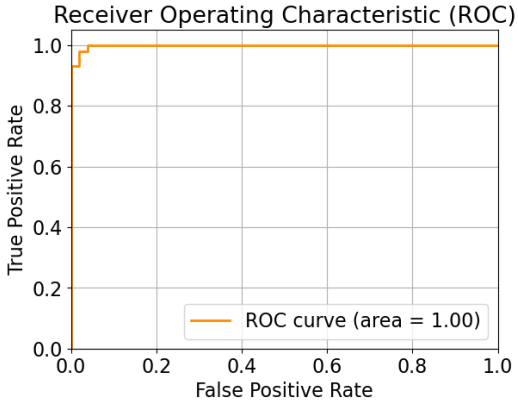


Figure 13. ROC curve for ViT.

The training and validation loss functions for DenseNet and ViT are demonstrated in Figures 14 and 15, respectively. These figures showcase the learning progression of each model throughout the training and validation phases. It is important to note that for ViT, there is a slight increase in the validation loss. This phenomenon can be attributed to the relatively small size of our dataset and the limited extent of the validation set. However, it is crucial to understand that this increase does not detrimentally affect the overall performance of the model, indicating that ViT maintains its robustness and effectiveness despite these constraints. This scenario underscores the importance of considering dataset characteristics when interpreting model behavior and performance metrics. It is worth noting that despite the high evaluation metrics achieved by both DenseNet and ViT models, instances of misclassification were observed, as shown in Figures 16 and 17. These misclassifications underscore the necessity of employing ensemble methods to enhance predictive performance.

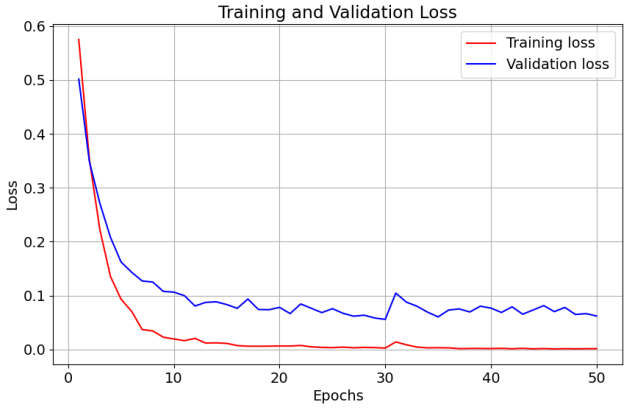


Figure 14. Loss functions for DenseNet.

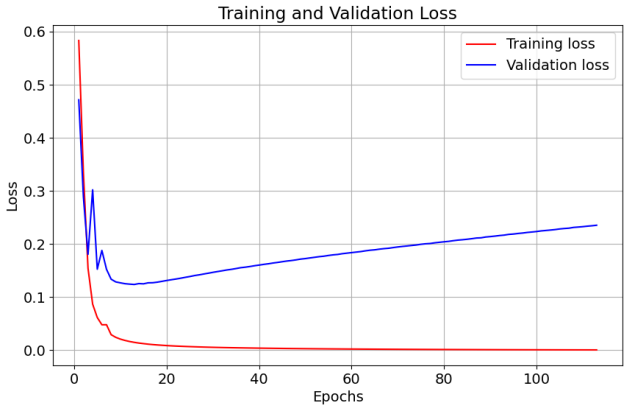


Figure 15. Loss functions for ViT.

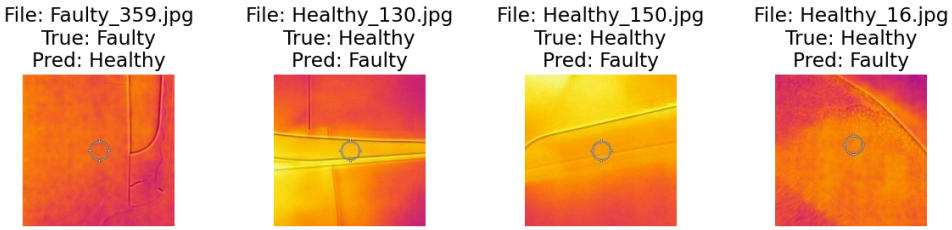


Figure 16. Misclassified images by DenseNet.

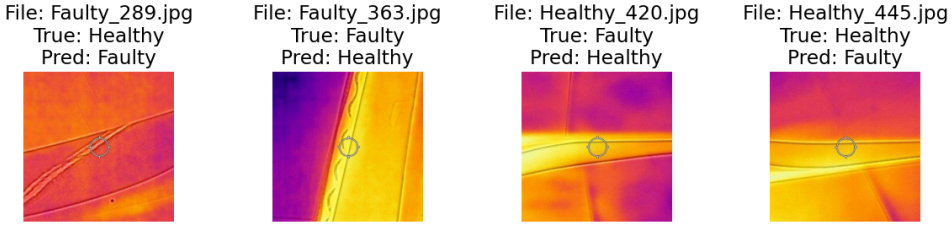


Figure 17. Misclassified images by ViT.

3.6.1. Implications for Ensemble Learning

The distinct misclassifications by different models suggest complementary strengths and weaknesses. This observation leads to the consideration of ensemble learning as a strategy to combine these strengths, potentially mitigating individual limitations and capitalizing on their collective capabilities for more accurate predictions. Ensemble learning, by combining models such as DenseNet and ViT, could exploit the complementarity of

DenseNet's densely connected features and ViT's ability to capture global dependencies, potentially leading to a more robust and accurate system for WTB inspection.

In summary, this analysis not only highlights the importance of model selection relative to the task and dataset characteristics but also underscores the capabilities of deep learning in image pattern recognition, even with limited data. It further emphasizes the potential of ensemble learning to enhance the accuracy and robustness of predictive models in practical applications.

3.6.2. Ensemble Learning

Ensemble learning is a machine learning approach where multiple models, often termed *weak learners*, are collaboratively employed to solve the same problem. The central concept is that the aggregation of predictions from several models can compensate for their individual weaknesses, thereby enhancing the accuracy and generalization on unseen data [73,74]. Common methods of ensemble learning include bagging, boosting, and stacking [75,76].

Bagging involves training numerous models in parallel on different subsets of the data to decrease variance and avert overfitting [75]. Boosting sequentially trains models, with each focusing on previously misclassified data points to minimize bias and improve robustness [75]. Stacking uses a new model to combine the predictions of multiple models, learning the optimal way to amalgamate these predictions [77].

As illustrated in Figure 18, the collective performance of multiple weak classifiers can surpass the effectiveness of individual classifiers, especially when each contributes a unique perspective [78]. This concept is central to our study, where we evaluate a variety of architectures such as DenseNet, ViTs, and ResNets. Each of these architectures exhibits distinct capabilities in classifying thermal images of WTBs [79]. The variability in misclassified samples among different models underscores the uniqueness of each classifier's approach to prediction, reinforcing the potential of an ensemble strategy to enhance overall accuracy [80].

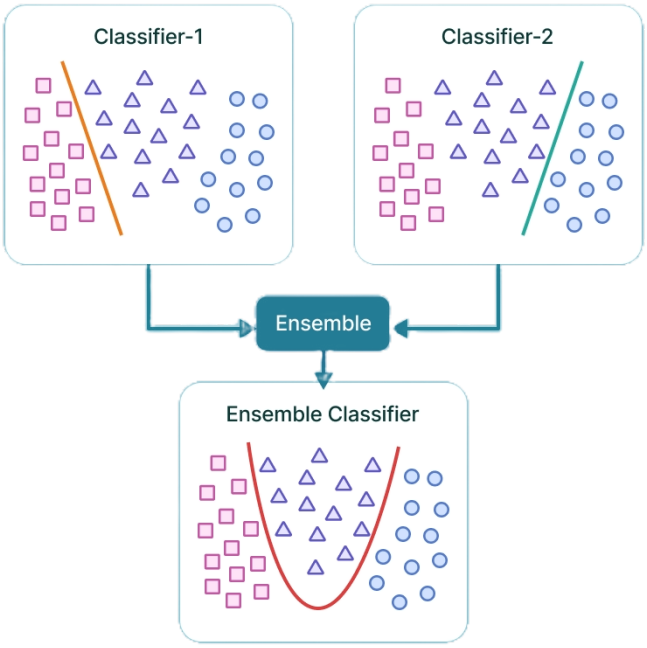


Figure 18. Illustration of ensemble learning with weak classifiers, which demonstrates the complementary strengths of individual models when combined [78].

Ensembling top-performing models with high accuracy and F1-scores can leverage their collective strengths and address individual misclassifications, thus improving the overall prediction performance [80]. For instance, combining models such as ViT and DenseNet161 is a strategic choice, given their already high accuracy and balance between precision and recall [80]. This ensemble approach can exploit DenseNet's dense convolu-

tional connections and ViT's global image analysis capabilities [80]. Practically, this could involve averaging the predicted probabilities from both models for each class, selecting the class with the highest combined probability as the final prediction. Alternatively, a more sophisticated meta-model could be developed to perform this aggregation, potentially incorporating additional promising models [80].

3.6.3. Ensemble Learning Results

The ensemble model, integrating the strengths of both ViT and DenseNet161, exhibited remarkable performance in detecting defects in WTBs. This model achieved a perfect classification accuracy of 100%, a testament to its efficiency and reliability in this high-stakes application. Attaining an accuracy of 100% on our dataset, the ensemble model surpassed the individual performances of its constituent models. As shown in Figure 19, while ViT and DenseNet161 individually achieved an accuracy of around 98%, their combined prowess in the ensemble configuration led to flawless classification accuracy.

To address the potential for misclassification, we conducted extensive testing of the ensemble model. Despite the individual models ViT and DenseNet161 showing instances of misclassification, when combined in the ensemble model, their strengths complemented each other, eliminating the occurrence of false positives and false negatives across multiple runs. Specifically, the ensemble model was run 10 times, each time achieving 100% accuracy with no observed misclassifications. This absence of errors can be attributed to the models' complementary nature in the ensemble, where one model's weaknesses are offset by the other's strengths, thus reinforcing the reliability of our approach for practical applications such as wind turbine maintenance.

In tuning our ensemble model, which combines the capabilities of ViT and DenseNet161, we meticulously explored a range of hyperparameters to ensure optimal performance. The process involved the following:

- **Model Weight Optimization**: We adjusted the relative contributions of ViT and DenseNet161 within the ensemble. Through extensive testing, we found a weight ratio of 60:40 to be most effective, though we evaluated ratios from an even split to a 70:30 distribution.
- **Integration Strategy**: Various integration strategies were tested to determine how best to combine the outputs of the two models. We experimented with simple averaging, weighted averaging (with weights from 0.1 to 0.9), and stacking approaches. A weighted averaging with a bias of 0.6 towards the ViT model's output was identified as the superior method.
- **Decision Thresholds**: To fine-tune our model's sensitivity and specificity, we employed Bayesian optimization techniques to find the optimal decision threshold for classifying images as defective or non-defective. After testing thresholds ranging from 0.5 to 0.8, we established 0.7 as the ideal balance point, which maximized both precision and recall.

These hyperparameter tuning steps were critical for harnessing the synergistic effects of the ensemble model, culminating in its exemplary defect detection performance as evidenced by our results.

Correspondingly, the ensemble model not only excelled in accuracy but also in other crucial evaluation metrics. The F1-score, which harmonizes precision and recall, reached the optimal value of 100%. This is a direct implication of the model achieving 100% accuracy in a balanced dataset, which inherently signifies no false positives or false negatives, thus leading to perfect precision and recall scores. In other words, the ensemble model's precision and recall were also at their theoretical maximum, demonstrating its exceptional capability in correctly identifying all defective and non-defective segments in the turbine blades without any errors. This level of performance in precision and recall, alongside the perfect F1-score, underscores the model's robustness beyond just accuracy. It highlights the model's unparalleled proficiency in both detecting defects (precision) and accurately classifying non-defective areas (recall), which is crucial in applications such as

wind turbine maintenance where the cost of misclassification can be high. Throughout our study, all evaluation metrics, including accuracy, precision, recall, and F1-score, have been comprehensively discussed for all models. The focus on the ensemble model's 100% accuracy underscores its comprehensive performance across all these metrics, reinforcing its suitability and reliability for high-precision applications like WTB defect detection.

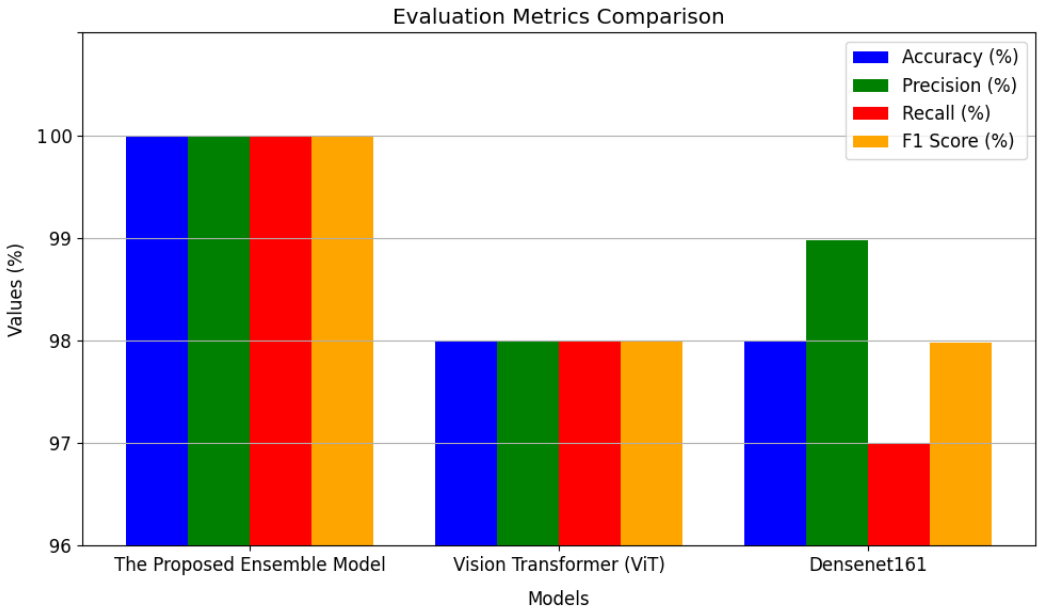


Figure 19. Comparison of accuracy and F1-Scores for ViT, DenseNet161, and the proposed implemented ensemble model.

To provide a comprehensive understanding of the model's performance, we compared the ensemble model with the individual performances of ViT and DenseNet161. As shown in Figure 19, the ensemble model not only improved accuracy but also enhanced precision, recall, and F1-scores compared to the individual models. The F1-score, a crucial metric in anomaly detection due to its balance of precision and recall, was particularly improved, reaching the optimal value of 100%. This demonstrates the ensemble model's advanced capability to correctly identify all defective and non-defective segments without error, which is essential for practical applications such as wind turbine maintenance where misclassification has significant consequences.

The ensemble model's success is attributed to the synergistic combination of ViT and DenseNet161. While ViT excels in capturing global features of the blade images, DenseNet161 is adept at recognizing finer, localized details. This complementary pairing allowed the ensemble model to leverage the global contextual understanding of ViT and the detailed analytical capability of DenseNet161, resulting in unparalleled classification accuracy.

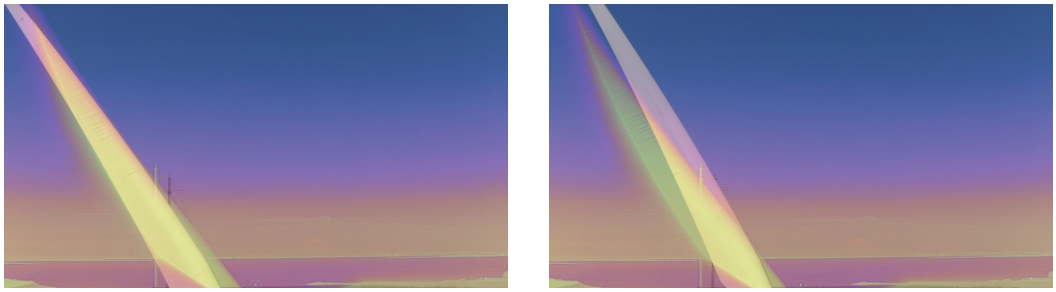
The perfect accuracy of the ensemble model has significant practical implications for wind turbine maintenance. It ensures highly accurate and reliable defect detection, paving the way for more effective predictive maintenance strategies. This could lead to reduced downtime, enhanced turbine longevity, and overall more efficient energy production.

3.6.4. External Validation on Operational Wind Turbine Data

To further assess the generalizability of our ensemble model, we conducted external validation using a unique dataset provided by Chen, Xiao [81]. This dataset contains drone-based optical and thermal imagery of wind turbine blades in operation, captured separately due to the technical constraints of using two distinct cameras that do not capture images simultaneously.

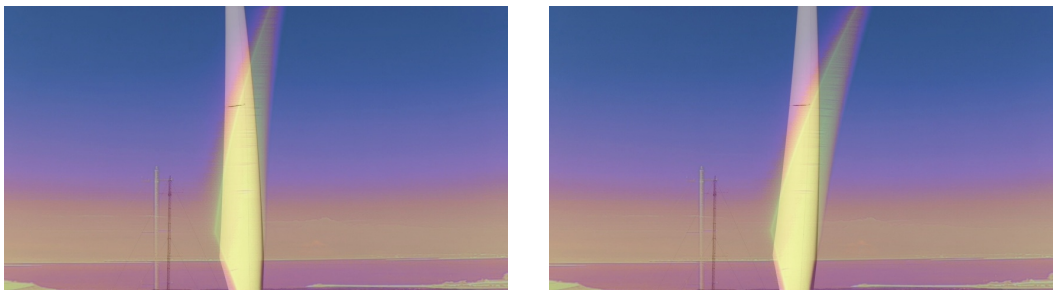
For this validation, we innovatively processed 100 healthy and 100 faulty images to create a fused dataset. The fusion process involved using OpenCV's alpha blending

technique, where we manipulated the alpha parameter to control the blend of optical and thermal images. Specifically, we set alpha to 0.7, allowing the optical image to contribute 70% to the fused image, with the remaining 30% contributed by the thermal image. This balanced approach was critical for overcoming the challenge of temporal misalignment, a prevalent issue with separately captured images. The alpha parameter's role was to ensure that the final fused image maintained a dominant visual influence from the optical data while effectively incorporating thermal data characteristics. Figures 20 and 21 showcase examples of these fused images, illustrating both healthy and faulty conditions.



(**a**) (**b**)

Figure 20. Examples of fused images from the external dataset showing a healthy wind turbine blade [81]. (**a**) Healthy blade—view 1; (**b**) Healthy blade—view 2.



(**a**) (**b**)

Figure 21. Examples of fused images from the external dataset showing a faulty wind turbine blade [81]. (**a**) Faulty blade—view 1; (**b**) Faulty blade—view 2.

Importantly, our pre-trained ensemble model, which integrates ViT and DenseNet161, was directly applied to this externally fused dataset without any retraining or fine-tuning. This approach was intended to rigorously test the model's ability to generalize and adapt to new, real-world data, a vital aspect of its practical applicability.

Despite the challenges in image fusion due to temporal misalignment, our ensemble model successfully classified all instances with 100% accuracy. This remarkable achievement in the external validation emphasizes the robustness and adaptability of our model to operational environments and enhances confidence in its deployment for wind turbine maintenance.

4. Conclusions and Future Directions

This research, conducted at Utah Valley University's Machine Learning and Drone Lab, represents a substantial breakthrough in the inspection of WTBs. Utilizing the FLIR Thermal Camera's Multi-Spectral Dynamic Imaging (MSX) technology, we created a unique dataset comprising 1000 thermal images. This dataset includes a comprehensive range of conditions, capturing both healthy and faulty blades with various forms of damage including cracks, holes, and erosion. The integration of high-resolution RGB imagery with thermal data using MSX technology enriches the dataset with precise visual cues, crucial for accurate anomaly detection and categorization.

Our rigorous testing and analysis process included a comprehensive evaluation of 35 different models, such as various versions of ResNet, VGG, MobileNet, along with ensemble models, ViT, and DenseNet161. This extensive evaluation highlights the importance of a holistic approach in machine learning model selection and optimization, especially for complex diagnostic tasks. The ensemble model emerged as the top performer, excelling in accuracy, precision, recall, and F1-score.

The proposed ensemble model, combining ViT and DenseNet161, leverages this dataset to achieve a 100% accuracy rate in defect detection, setting a new paradigm in renewable energy infrastructure maintenance. This model exemplifies the potential of advanced machine learning techniques in complex industrial settings. The effectiveness of the ensemble model is attributed to the synergistic blend of ViT's global feature analysis and DenseNet161's local pattern recognition capabilities, which, along with the meticulously curated dataset, allows for the nuanced and highly accurate identification of defects.

Looking forward, the research opens up several avenues for future exploration. Expanding the dataset will enhance the model's robustness and improve diagnostic accuracy across diverse scenarios. Advancements in defect detection methodologies, particularly in the classification of defect types, are vital for more detailed diagnostics. Exploring the depth assessment of defects could offer comprehensive insights into their severity. Furthermore, integrating thermal imagery with data from other sensors, such as acoustic and vibration monitors, proposes a multifaceted diagnostic approach. The use of drones equipped with thermal cameras for data collection offers an efficient method for comprehensive inspections in challenging environments.

In summary, this study lays a solid foundation for future research in WTB inspection. It emphasizes the potential of integrating advanced imaging technologies and machine learning models to develop more accurate, efficient, and practical maintenance tools in the renewable energy sector. Future research directions include the following:

- Expanding the dataset to enhance model robustness and diagnostic accuracy across diverse scenarios.
- Advancing defect detection methodologies to classify types of defects for more detailed diagnostics.
- Exploring the depth assessment of defects to provide comprehensive insights into their severity.
- Integrating thermal imagery with data from other sensors, such as acoustic and vibration monitors, for a multifaceted diagnostic approach.
- Implementing drones equipped with thermal cameras for efficient and comprehensive data collection in challenging environments.

Author Contributions: M.M. conceptualized and designed the methodology, developed the software, and prepared the original draft of the manuscript. M.S. was responsible for funding, conceptualization, data curation, formal analysis, and contributed to the manuscript through review and editing, also providing supervision. M.A.S.M. and A.C.S. assisted in reviewing and editing the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Utah System of Higher Education (USHE)-Deep Talent Technology Initiative Grant 20210016UT.

Data Availability Statement: The created dataset can be found at https://github.com/MoShekaramiz/Small-WTB-Thermal1.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

Adam	Adaptive Moment Estimation
AlexNet	Alex Neural Network
ANN	Artificial Neural Network

AUC	Area Under the Curve
CNN	Convolutional Neural Network
FPR	False Positive Rate
IoT	Internet of Thing
ReLU	Rectified Linear Unit
ResNet	Residual Neural Network
RGB	Red, Green, Blue
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
TPR	True Positive Rate
VGG	Visual Geometry Group
ViT	Vision Transformer
WTB	Wind Turbine Blade
Xception	Extreme Inception

References

1. Maes, W.H.; Huete, A.R.; Steppe, K. Optimizing the processing of UAV-based thermal imagery. *Remote Sens.* **2017**, *9*, 476. [CrossRef]
2. Martin, R.; Sabato, A.; Schoenberg, A.; Giles, R.; Niezrecki, C. Comparison of Nondestructive Testing Techniques for the Inspection Of Wind Turbine Blades' Spar Caps. *Wind. Energy* **2018**, *21*, 980–996. [CrossRef]
3. Wang, G.; Li, C.; Ma, Y.; Zheng, A.; Tang, J.; Luo, B. RGB-T Saliency Detection Benchmark: Dataset, Baselines, Analysis and a Novel Approach. In Proceedings of the Image and Graphics Technologies and Applications, Beijing, China, 8–10 April 2018; pp. 359–369.
4. Sanati, H.; Wood, D.; Sun, Q. Condition monitoring of wind turbine blades using active and passive thermography. *Appl. Sci.* **2018**, *8*, 2004. [CrossRef]
5. Zhang, Q.; Huang, N.; Yang, L.; Zhang, D.; Shan, C.; Han, J. RGB-T Salient object detection via fusing multi-level CNN features. *IEEE Trans. Image Process.* **2020**, *29*, 3321–3335. [CrossRef] [PubMed]
6. Chaudhuri, S.; Stamm, M.; Krankenhagen, R. Weather-dependent passive thermography and thermal simulation of in-service wind turbine blades. *J. Phys. Conf. Ser.* **2023**, *2507*, 012025. [CrossRef]
7. Zhou, W.; Wang, Z.; Zhang, M.; Wang, L. Wind Turbine Actual Defects Detection Based on Visible and Infrared Image Fusion. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1–8. [CrossRef]
8. Zhu, J.; Wen, C.; Liu, J. Defect identification of wind turbine blade based on multi-feature fusion residual network and transfer learning. *Energy Sci. Eng.* **2022**, *10*, 219–229. [CrossRef]
9. Kwon, K.A.; Choi, M.Y.; Park, H.S.; Park, J.H.; Huh, Y.H.; Choi, W.J. Quantitative defects detection in wind turbine blade using optical infrared thermography. *J. Korean Soc. Nondestruct. Test.* **2015**, *35*, 25–30. [CrossRef]
10. Manohar, A.; Tippmann, J.; di Scalea, F.L. Localization of defects in wind turbine blades and defect depth estimation using infrared thermography. In Proceedings of the Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems, San Diego, CA, USA, 11–15 March 2012; Volume 8345, pp. 444–460.
11. Ligocki, A.; Jelinek, A.; Zalud, L.; Rahtu, E. Fully automated DCNN-based thermal images annotation using neural network pretrained on RGB data. *Sensors* **2021**, *21*, 1552. [CrossRef] [PubMed]
12. Liu, J.; Zhang, S.; Wang, S.; Metaxas, D.N. Multispectral deep neural networks for pedestrian detection. *arXiv* **2016**, arXiv:1611.02644.
13. Liu, H.; Chen, F.; Zeng, Z.; Tan, X. AMFuse: Add–Multiply-Based Cross-Modal Fusion Network for Multi-Spectral Semantic Segmentation. *Remote Sens.* **2022**, *14*, 3368. [CrossRef]
14. Shopovska, I.; Jovanov, L.; Philips, W. Deep visible and thermal image fusion for enhanced pedestrian visibility. *Sensors* **2019**, *19*, 3727. [CrossRef] [PubMed]
15. French, G.; Finlayson, G.; Mackiewicz, M. Multi-spectral pedestrian detection via image fusion and deep neural networks. *J. Imaging Sci. Technol.* **2018**, 176–181.
16. Gallagher, J.E.; Oughton, E.J. Assessing thermal imagery integration into object detection methods on air-based collection platforms. *Sci. Rep.* **2023**, *13*, 8491. [CrossRef] [PubMed]
17. Krizhevsky, A.; Sutskever, I.; Hinton, G. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
18. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 618–626.
19. Wang, L. AI-powered drone-based automated inspection of fast. *Light. Sci. Appl.* **2023**, *12*, 63. [CrossRef]
20. Thenmozhi, R.; Amutha, B.; Valsalakumar, S.; Rajagopal, T.; Sundaram, S. Application of MSVPC- 5G multicast SDN network eminence video transmission in drone thermal imaging for solar farm monitoring. *Energies* **2021**, *14*, 8255.
21. Feng, Q.; Li, R.; Nie, B.; Liu, S.; Zhao, L.; Zhang, H. Literature review: Theory and application of in-line inspection technologies for oil and gas pipeline girth weld defection. *Sensors* **2016**, *17*, 50. [CrossRef]

22. Lee, S.; Ahn, H.; Seo, J.; Chung, Y.; Park, J.; Pan, S. Practical monitoring of undergrown pigs for IoT-based large-scale smart farm. *IEEE Access* **2019**, *7*, 173796–173810. [CrossRef]
23. Jiang, W.; Guo, Z.; Zhang, H.; Cheng, L.; Tian, Y. Target-aware deep feature compression for power intelligent inspection tracking. *J. Electr. Comput. Eng.* **2022**, *2022*, 3161551. [CrossRef]
24. Ravishankar, P.; Hwang, S.; Zhang, J.; Khalilullah, I.X.; Eren-Tokgoz, B. Darts-drone and artificial intelligence reconsolidated technological solution for increasing the oil and gas pipeline resilience. *Int. J. Disaster Risk Sci.* **2022**, *13*, 810–821. [CrossRef]
25. Wu, X.; Lv, L.; Yao, Z.; Wang, E.; Ren, X.; Pang, R.; Wang, H.; Zhang, Y. Efficient and accurate damage detector for wind turbine blade images. *IEEE Access* **2022**, *10*, 123378–123386.
26. Zhang, C.; Yang, T.; Yang, J. Image recognition of wind turbine blade defects using attention-based MobileNetv1-YOLOv4 and transfer learning. *Sensors* **2022**, *22*, 6009. [CrossRef]
27. Peng, Y.; Tang, Z.; Zhao, G.; Cao, G.; Wu, C. Motion blur removal for UAV-based wind turbine blade images using synthetic datasets. *Remote Sens.* **2021**, *14*, 87. [CrossRef]
28. Zhang, R.; Wen, C. SOD-YOLO: A Small Target Defect Detection Algorithm for Wind Turbine Blades Based on Improved YOLOv5. *Adv. Theory Simulations* **2022**, *5*, 2100631. [CrossRef]
29. Denhof, D.; Staar, B.; Lütjen, M.; Freitag, M. Automatic optical surface inspection of wind turbine rotor blades using convolutional neural networks. *Procedia CIRP* **2019**, *81*, 1166–1170. [CrossRef]
30. Qiu, Z.; Wang, S.; Zeng, Z.; Yu, D. Automatic visual defects inspection of wind turbine blades via YOLO-based small object detection approach. *J. Electron. Imaging* **2019**, *28*, 043023. [CrossRef]
31. Shaheed, H.H.; Aggarwal, R. Wind Turbine Surface Defect Detection Analysis from UAVs Using U-Net Architecture. In Proceedings of the Science and Information Conference, Grand Rapids, MI, USA, 13–14 July 2022; pp. 499–511.
32. Carnero, A.; Martín, C.; Díaz, M. Portable motorized telescope system for wind turbine blades damage detection. *Eng. Rep.* **2023**, e12618. [CrossRef]
33. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26–30 June 2016; pp. 770–778.
34. Zehui, L.; Liu, P.; Huang, L.; Chen, J.; Qiu, X.; Huang, X. Dropattention: A regularization method for fully-connected self-attention networks. *arXiv* **2019**, arXiv:1907.11065.
35. Hou, S.; Wang, Z. Weighted channel dropout for regularization of deep convolutional neural network. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–12 February 2019; Volume 33, pp. 8425–8432.
36. Zhang, A.; Lipton, Z.C.; Li, M.; Smola, A.J. *Dive into Deep Learning*; Cambridge University Press: Cambridge, UK , 2023.
37. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
38. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, San Diego, CA, USA, 15–17 September 2015; pp. 448–456.
39. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity mappings in deep residual networks. In Proceedings of the Computer Vision–ECCV: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; pp. 630–645.
40. Wang, C.-F. A Basic Introduction to Separable Convolutions. 2018. Available online: https://towardsdatascience.com/a-basic-introduction-to-separable-convolutions-b99ec3102728 (accessed on 30 January 2024).
41. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
42. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50× fewer parameters and <0.5 MB model size. *arXiv* **2016**, arXiv:1602.07360.
43. Shafiee, M.J.; Li, F.; Chwyl, B.; Wong, A. SquishedNets: Squishing SqueezeNet further for edge device scenarios via deep evolutionary synthesis. *arXiv* **2017**, arXiv:1711.07459.
44. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
45. Jiang, C.; Jiang, C.; Chen, D.; Hu, F. Densely connected neural networks for nonlinear regression. *Entropy* **2022**, *24*, 876. [CrossRef]
46. Zhang, J.; Zhao, J.; Lin, H.; Tan, Y.; Cheng, J. High-speed chemical imaging by Dense-Net learning of femtosecond stimulated raman scattering. *J. Phys. Chem. Lett.* **2020**, *11*, 8573–8578. [CrossRef] [PubMed]
47. Baoyuan, C.; Shen, Y.; Sun, K. Research on object detection algorithm based on multilayer information fusion. *Math. Probl. Eng.* **2020**, *2020*, 9076857.
48. Niyongabo, J.; Zhang, Y.; Ndikumagenge, J. Bearing fault detection and diagnosis based on densely connected convolutional networks. *Acta Mech. Autom.* **2022**, *16*, 130–135. [CrossRef]
49. Huang, G.; Liu, Z.; Pleiss, G.; Maaten, L.; Weinberger, K. Convolutional networks with dense connectivity. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 8704–8716. [CrossRef]
50. Wang, Y.; Li, H.; Jia, P.; Zhang, G.; Wang, T.; Hao, X. Multi-scale DenseNet-based aircraft detection from remote sensing images. *Sensors* **2019**, *19*, 5270. [CrossRef] [PubMed]
51. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.

52. Aufar, Y.; Abdillah, M.; Romadoni, J. Web-based CNN application for Arabica coffee leaf disease prediction in smart agriculture. *J. RESTI Rekayasa Sist. Dan Teknol. Inf.* **2023**, *7*, 71–79.
53. Cheng, A.C.; Lin, C.H.; Juan, D.C.; Wei, W.; Sun, M. InstaNAS: Instance-aware neural architecture search. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 3577–3584.
54. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16 × 16 words: Transformers for image recognition at scale. *arXiv* **2021**, arXiv:2010.11929.
55. Zhou, D.; Shi, Y.; Kang, B.; Yu, W.; Jiang, Z.; Li, Y.; Jin, X.; Hou, Q.; Feng, J. Refiner: Refining self-attention for vision transformers. *arXiv* **2021**, arXiv:2106.03714.
56. Yuan, L.; Chen, Y.; Wang, T.; Yu, W.; Shi, Y.; Jiang, Z.H.; Tay, F.E.; Feng, J.; Yan, S. Tokens-to-token ViT: Training vision transformers from scratch on ImageNet. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, BC, Canada, 11–17 October 2021; pp. 558–567.
57. Chu, X.; Tian, Z.; Zhang, B.; Wang, X.; Wei, X.; Xia, H.; Shen, C. Conditional positional encodings for vision transformers. *arXiv* **2021**, arXiv:2102.10882
58. Xia, R.; Wang, J.; Xue, C.; Deng, B.; Wang, F. EIT: Efficiently Lead Inductive Biases to ViT. *arXiv* **2022**, arXiv:2203.07116v1.
59. Zhang, X.; Tian, Y.; Huang, W.; Ye, Q.; Dai, Q.; Xie, L.; Tian, Q. HIViT: Hierarchical vision transformer meets masked image modeling. *arXiv* **2022**, arXiv:2205.14949.
60. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
61. Liu, L.; Jiang, X.; Saerbeck, M.; Dauwels, J. Recurrent affine transform encoder for image representation. *IEEE Access* **2022**, *10*, 18653–18666. [CrossRef]
62. FLIR Systems. What Is MSX®? 2023. Available online: www.flir.com/discover/professional-tools/what-is-msx (accessed on 19 November 2023).
63. FLIR Systems. FLIR C5 Compact Thermal Camera. 2023. Available online: https://www.flir.com/products/c5 (accessed on 19 November 2023).
64. Usamentiaga, R.; Venegas, P.; Guerediaga, J.; Vega, L.; Molleda, J.; Bulnes, F.G. Infrared thermography for temperature measurement and non-destructive testing. *Sensors* **2014**, *14*, 12305–12348. [CrossRef] [PubMed]
65. Shi, Z.; Zhao, Y.; Liu, Z.; Zhang, Y.; Ma, L. Diagnosis and classification decision analysis of overheating defects of substation equipment based on infrared detection technology. *Sci. Program.* **2021**, *2021*, 3356044. [CrossRef]
66. Aguerre, J.P.; García-Nevado, E.; Acuña Paz y Miño, J.; Fernández, E.; Beckers, B. Physically based simulation and rendering of urban thermography. *Comput. Graph. Forum* **2020**, *39*, 377–391. [CrossRef]
67. Zhao, H.; Ji, Z.; Li, N.; Gu, J.; Li, Y. Target detection over the diurnal cycle using a multispectral infrared sensor. *Sensors* **2016**, *17*, 56. [CrossRef]
68. Wang, P.; Wang, S.; Zhang, Y.; Duan, X. Multispectral Image under Tissue Classification Algorithm in Screening of Cervical Cancer. *J. Healthc. Eng.* **2022**, *2022*, 9048123. [CrossRef]
69. Mehl, P.; Chao, K.; Kim, M.; Chen, Y. Detection of defects on selected apple cultivars using hyperspectral and multispectral image analysis. *Appl. Eng. Agric.* **2002**, *18*, 219.
70. Khan, H.A.; Thomas, J.B.; Hardeberg, J.Y.; Laligant, O. Illuminant estimation in multispectral imaging. *J. Opt. Soc. Am. JOSA* **2017**, *34*, 1085–1098. [CrossRef] [PubMed]
71. Martínez, M.Á.; Etchebehere, S.; Valero, E.M.; Nieves, J.L. Improving unsupervised saliency detection by migrating from RGB to multispectral images. *Color Res. Appl.* **2019**, *44*, 875–885. [CrossRef]
72. Weiss, K.; Khoshgoftaar, T.M.; Wang, D. A survey of transfer learning. *J. Big Data* **2016**, *3*, 1–40. [CrossRef]
73. Nguyen, K.; Chen, W.; Lin, B.; Seeboonruang, U. Comparison of ensemble machine learning methods for soil erosion pin measurements. *ISPRS Int. J. Geo Inf.* **2021**, *10*, 42. [CrossRef]
74. Jain, A.; Duin, R.; Mao, J. Statistical pattern recognition: A review. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 4–37. [CrossRef]
75. Opitz, D.; Maclin, R. Popular ensemble methods: An empirical study. *J. Artif. Intell. Res.* **1999**, *11*, 169–198. [CrossRef]
76. Moradzadeh, A.; Mansour-Saatloo, A.; Mohammadi-Ivatloo, B.; Anvari-Moghaddam, A. Performance evaluation of two machine learning techniques in heating and cooling loads forecasting of residential buildings. *Appl. Sci.* **2020**, *10*, 3829. [CrossRef]
77. Wang, M.; Zhang, Y.; Qin, C.; Liu, P.; Zhang, Q. Option pricing model combining ensemble learning methods and network learning structure. *Math. Probl. Eng.* **2022**, *2022*, 2590940. [CrossRef]
78. Kundu, R.; Singh, P.K.; Ferrara, M.; Ahmadian, A.; Sarkar, R. ET-NET: An ensemble of transfer learning models for prediction of COVID-19 infection through chest CT-scan images. *Multimed. Tools Appl.* **2022**, *81*, 31–50. [CrossRef] [PubMed]
79. Divina, F.; Gilson, A.; Gómez-Vela, F.; Torres, M.; Torres, J. Stacking ensemble learning for short-term electricity consumption forecasting. *Energies* **2018**, *11*, 949. [CrossRef]
80. Rokach, L. Ensemble-based classifiers. *Artif. Intell. Rev.* **2009**, *33*, 1–39. [CrossRef]
81. Chen, X. Technical University of Denmark. Drone-based optical and thermal videos of rotor blades taken in normal wind turbine operation. *IEEE Dataport* **2023**. [CrossRef]