Check for updates

Deep learning model for detection of hotspots using infrared thermographic images of electrical installations

Ezechukwu Kalu Ukiwe[1*], Steve A. Adeshina[2], Tsado Jacob[3] and Bukola Babatunde Adetokun[1]

*Correspondence:
ezeukiwe@yahoo.com

[1] Electrical/Electronic
Engineering Department, Nile
University of Nigeria, Abuja,
Nigeria
[2] Computer Engineering
Department, Nile University
of Nigeria, Abuja, Nigeria
[3] Electrical/Electronic
Engineering Department, Federal
University of Technology Minna,
Minna, Nigeria

Abstract

Hotspots in electrical power equipment or installations are a major issue whenever it occurs within the power system. Factors responsible for this phenomenon are many, sometimes inter-related and other times they are isolated. Electrical hotspots caused by poor connections are common. Deep learning models have become popular for diagnosing anomalies in physical and biological systems, by the instrumentality of feature extraction of images in convolutional neural networks. In this work, a VGG-16 deep neural network model is applied for identifying electrical hotspots by means of transfer learning. This model was achieved by first augmenting the acquired infrared thermographic images, using the pre-trained ImageNet weights of the VGG-16 algorithm with additional global average pooling in place of conventional fully connected layers and a softmax layer at the output. With the categorical cross-entropy loss function, the model was implemented using the Adam optimizer at learning rate of 0.0001 as well as some variants of the Adam optimization algorithm. On evaluation, with a test IRT image dataset, and a comparison with similar works, the research showed that a better accuracy of 99.98% in identification of electrical hotspots was achieved. The model shows good score in performance metrics like accuracy, precision, recall, and F_1-score. The obtained results proved the potential of deep learning using computer vision parameters for infrared thermographic identification of electrical hotspots in power system installations. Also, there is need for careful selection of the IR sensor's thermal range during image acquisition, and suitable choice of color palette would make for easy hotspot isolation, reduce the pixel to pixel temperature differential across any of the images, and easily highlight the critical region of interest with high pixel values. However, it makes edge detection difficult for human visual perception which computer vision-based deep learning model could overcome.

Keywords: Electrical hotspot detection, Deep learning, VGG-16, Adam optimizer, Condition monitoring, Infrared inspection, Power system monitoring

Introduction

Electrical hotspot is a major issue in electrical substations whenever they occur. It can lead to inefficient power flow, power outage, and even serious accident. Hotspot faults tends to degenerate with time. Like most problems follows the basic concept as recalled by Adedigba et al. [1] that early problem identification is vital to limit its further

Ukiwe *et al. Journal of Electrical Systems and Inf Technol*   (2024) 11:24

Page 2 of 25

degeneration, enhance adequate intervention program, improve chances of success for such intervention, and reduce rate of permanent damage.

Hotspot faults in electrical systems as noted by Oluseyi et al. [2] can result from many factors, such as poor connections, corroded joints, overload, mechanical vibrations, insulation failure, and poor cooling system. The effect is mainly severe at high voltage high current installations where special procedure is required to address the issue. In the works of Sousa et al. [3], Usamentiaga et al. [4], Aidossov et al. [5], and Ng [6], the authors observed that infrared (IR) thermal imaging systems are one of the most effective means of detecting hotspots in electrical power installations because of their non-invasive nature, low risk to personnel, ease of use, ability to produce a good thermal distribution of the equipment in service, cost efficiency, etc. However, Epperly et al. [7] inferred that they require expert power system analysts to interpret the images and for best results.

According to Balakrishnan et al. [8], computer vision-based deep learning applications are being developed to aid the system experts in the interpretation of the thermal images. Ukiwe et al. [9] perceived that efficient imaging systems provides the bedrock for effective computer vision applications that aids human operators. The human visual system may not accurately discriminate the boundaries of equipment shown by their thermal images and the pixel-wise progression of heat profile across the equipment. So, the process of feature extraction using deep convolutional neural network is imperative, and a lot of information can be acquired from the pixel values of the IR images, as shown by Gao et al. [10].

Choi et al. [11] saw that deep learning-based applications are better suited to operations with images because they apply convolutional operation on patches of the image for inputs to certain nodes of the model with recognition of spatial correlation between pixels in the images than artificial neural networks that usually takes each pixel as a separate input to all nodes in the input layer of the model and have been found to be adaptable to different optimization algorithms and easy to implement on wide variety of linear and nonlinear problems, according to Ahmed et al. [12]. Hotspots can be detected, prevented, and removed, if found early. IR thermography (IRT)-based method is the standard method to detect it in operational electrical equipment. To observe the progression of the hotspots and flag them as an anomaly in certain equipment would need knowledge of the power system. System analysts without requisite experience would have problem flagging issues at the installations. We are therefore motivated to apply a computer vision-based deep learning model because of the flexibility and numerous options it affords in terms of different features or architectures therein as shown by Wang et al. [13] that it would assist power system operators for identifying hotspots as anomalies in electrical installations before they could lead to catastrophic incident. Balla et al. [14] pointed out that the deep learning model can be integrated into SCADA systems for requisite annunciation, control, and protection actions.

Most computer vision applications often incorporate practical units for image acquisition, image preprocessing, image segmentation, feature extraction, classification, or detection as noted in the works of Soni [15], Wiley and Lucas [16], Rybchak and Basystiuk [17]. And the major issue as seen by Acquaah [18] is how to extract the features from thermographic images using suitable model to capture the extracted features in

Ukiwe *et al. Journal of Electrical Systems and Inf Technol*　　(2024) 11:24

Page 3 of 25

order to identify the area of defect. The challenge lies in this process, in the sense that adopting a good model to pinpoint important information from the thermal images could require mathematical analysis for optimization of any applied activation, objective or loss functions, according to Mustapha et al. [19]. Sometimes, the choice is made through rigorous experimentation or as presented by [20] through automatic tuning of requisite hyperparameters.

In this research, we have implemented a VGG-16 network using transfer learning approach for recognizing hotspots in electrical power system. A pre-trained VGG-16 network as an expert learning model is trained to identify anomalies in IRT images of power installations. VGG-16 is a deep learning model useful for many object classification and detection models and used in this work for hotspot detection as anomalies in IR images. The VGG-16 model was trained using digital IRT images of equipment in power substations and obtained 99.98% accuracy. The result shows the prospect of the computer vision-based model for identification electrical hotspots in the field electrical installation inspections, and on general electrical power system analysis. Prospect exists for further research to determine the root causes of the anomaly based on knowledge of thermal signature of the equipment where the hotspot occurs.

Rest of the paper gives a detailed explanation of the process. Section two gives insight on the reviewed literature in relation to the research. The third section discusses the theoretical concepts deplored for implementing the deep learning model. It succeeding section describes the overall design of the proposed hotspot detection system. And the last part before the conclusion provides details of the results of experiments conducted with associated evaluation of each applicable model.

Literature review

As research into the area of infrared imaging matures, authors widens their works to accommodate and apply new techniques of image processing. Hence, computer vision and deep learning tools become the choice platform to enhance the research work. This can be attributed to the different and numerous ways deep learning networks can be implemented, as well as creating building blocks for further research.

Review of anomaly detection in infrared thermography using deep learning and other novel techniques

The field of infrared technology for condition monitoring of electrical installation can be challenging because the thermal images in most cases are not easy to interpret and they come loaded with lots of information which can be decomposed using different image transformation techniques. Hotspot detection is one area that has benefitted from these novel image processing methods and convolutional neural networks have become invaluable tools for extracting such critical features of interests from the images. The convolutional layers are the major building blocks of deep learning models which can be used to detect anomalies in diverse areas not limited to electrical engineering. For instance, Yang et al. [21] developed a deep learning model to that is capable of detecting cracks in steel elements after excitation with an external heat source. The temperature variation regions of normal and abnormal state were observed and compared. Faster Region-based CNN (Faster-RCNN) was applied with a test accuracy of 95.54% and a

mean average precision (mAP) of 92.41%, whereas Ding et al. [22] presented a technique for identifying hotspot in a dataset characterized by a dynamic non-homogenous space, such as a road traffic using a novel ensemble learning approach. Usually, the investigations are conducted in a practical environment, however Fang et al. [23] designed an automatic dep learning defect detector using opto-exciter to identify anomalies in plexiglass, steel, and carbon-fiber-reinforced material in a simulated environment using YOLO-V3, Faster RCNN, center-mask, U-net, Res-U-net, and Mask RCNN. Choudhary et al. [24] discovered that different defects in rotary parts of an electric machine can be diagnosed using LeNet-5 which the authors found to be better than Artificial Neural Network (ANN). Infrared sensors are excellent surface temperature detecting devices and hence can be deployed to monitor temperature profile across the periphery of equipment. Thus, Das et al. [25] developed a metal-oxide surge arrester surface contamination detector using ResNet50 model that was trained with IR images through transfer learning. Chandra et al. [26] proposed a DL method for analyzing concrete pavement structures considering seasonal changes in weather and found the model yielding up to 96.47% accuracy, which the authors said outperformed a comparative surge arrester monitoring model based on ResNet. Janssens et al. [27] investigated the process of application of DL to IR thermal video for independent condition monitoring of electric machine with 95% accuracy for fault identification and 91.675 for oil level forecasting using VGG network. Fanchiang et al. [28] presented a real-time fault condition monitoring of dry-type transformer using Wasserstein Autoencoder Reconstruction in series with a differential image classifier (DIC) that is a light weight model when compared with other models like VGG-16, ResNet50, LeNet-5, SqueezeNet, ShuffleNet, and MobileNet v1. Jiang et al. [29] researched on faults in transformer bushings using a combination of Mask RCNN and improved PCNN with 98% accuracy. Fanchiang and Kuo [30] deployed a generative model based on Variational Autoencoder Generative Adversarial Network (VA-GAN) for identifying thermal anomaly in cast resin dry-type transformer with impressive performance with respect to accuracy, AUROC, F1-score and mean accuracy. Fang et al. [31] applied a partial supervised learning using synthesis based on GAN for hotspot fault detection in transformers. The author applied a feature extraction scheme that produced up to 82.2% accuracy for ordinary machine defect and 86.2% for overheating effect. Mlakic et al. [32] analyzed defects in 10/0.4kV distribution transformer using DL-based CNN model IR thermography. Jangblad [33] researched on the use of IR images for DCNN-based object detection network.

Jaffery and Dubey [34] developed a color dependent segmentation method to mark high temperature areas within thermograms of power equipment. The authors used a RGB and grayscale histograms of equipment during energized and off-line states to gauge the degree of increase or decrease in temperature of the. Kumar and Ansari [35] used a non-intrusive IR monitoring method to assess thermal profile of ZnO surge arrester using watershed transform for isolating features in images in order to identify link between heat profile of the equipment and its leakage current. Novizon et al. [36] proposed an artificial neural network (ANN)-based classification model for condition monitoring of metal-oxide surge arresters using heat profile of equipment together with the associated third harmonic component of its leakage current. Alvarado-Hernandez et al. [37] observed the faulty condition of rotary elements in an induction motor like the

Ukiwe *et al. Journal of Electrical Systems and Inf Technol*   (2024) 11:24

Page 5 of 25

bearing and gearing system using IR image acquisition system together with intelligent system based on deep neural network with accuracy of up to 96.1% when interfaced with Principal Component Analysis. Balakrishnan et al. [8] explained the basic process of IRT for monitoring state of electric equipment. The authors underscored and explained two common methods of hotspot detection to include quantitative approaches involving taking specific temperature profile of equipment, and qualitative methods entail a comparative thermal measurement of a particular point or area with respect to similar/comparable area under similar operating condition. Parashar et al. [38] inferred that IRT remains a valuable tool for detection or identification of faults/defects and would also serve for fault prediction and suggested that morphological reconstruction of images such as sharpening to remove noise therein, would aid improve the intelligent models. And, Liu et al. [39] noted the impact of human factor for recognizing faults in equipment from analysis of their IR images using smart systems based on feature extraction. The author pointed out that uncertainty in the ability of IR system to differentiate between various types of faults in the mechanical parts of rotary machines and presented an IRT-based CNN technique for tracing the type of rotary part defects like abrasion, loose bearing, shaft misalignment and combination of abrasion and misalignment, responsible for the manifested overheating parts with 95.8% accuracy.

Review of condition monitoring techniques for anomaly detection literature using VGG-16

Due to its ease of use and efficiency, as shown in the work of Simonyan and Zisserman [40], the VGG-16 network has been applied by researchers in solving most problems in different areas of life. Liu and Wang [41] proposed a power system diagnostic tool based on VGG-16 feature extraction of line graph data obtained from converted synchro-phase measured data of the system, whereas Younis et al. [42] proposed a VGG-16-based model for detecting anomalies in the magnetic resonance images (MRI) of the human brain occasioned by tumors therein. The method produced up to 98.5% accuracy better than a comparable Ensemble and CNN-based models. Piekarski et al. [43] interest in visible wideband radiation and far-field-IR absorption spectrometry led them to observe anomalies in high-powered synchrotron light beam caused by inefficient or ineffective optical sensors over a given period, using VGG-16-based neural network with 92% accuracy. Dang et al. [44] used VGG-16-based neural network to investigate faults in power transformers glimpsed from acoustic signals generated by the equipment during periods of normal and mal-operation using Gammatone Filters to derive the associated Gammatone Frequency Cepstral Coefficients (GFCC) that would be recognized by the VGG-16 network. The method was able to achieve 95% accuracy for different states of a test 10kV dry-type transformer. The VGG-16 network has been effective for detecting anomalies in monochrome images. Due to the public health issue associated with pneumonia, Sharma and Guleria [45] implemented a DL model based on VGG-16 to identify pneumonia with up to 95.4% accuracy, in a dataset containing anterior and posterior chest X-ray (CXR) monochromatic image of patients. Du et al. [46] showed that GoogleNet and VGG-16 can produce good classification accuracy for smart classification of Silicon PV Cell anomalies. Alatawi et al. [47] applied VGG-16 to identify diseases in plants by monitoring their leaf conditions with 95.2% accuracy on the test dataset. Similarly, Ahmad et al. [48] developed different classification and identification models

Ukiwe *et al. Journal of Electrical Systems and Inf Technol*   (2024) 11:24

Page 6 of 25

for tomato leaf diseases using four deep learning methodologies like ResNet, Inception V3, VGG-19, and VGG-16 for feature extraction as well as for tuning the hyperparameters of the networks and noted good performance across the deep learning networks even though the Inception V3 came tops with 99.6% accuracy.

In most cases, the VGG-16 algorithm is applied independently, but have proven to be just effective when used alongside other algorithms as Rasyid et al. [49] used VGG-16 with cosine similarity algorithm to develop an image-based electronic device recommender system to complement the traditional word-based systems. The model achieved up to 94.38% mean average precision in one of the test categories. Sheriff et al. [50] proposed a VGG-16-based web application model for diagnosing lung carcinoma to aid early detection and reduce error in diagnosis. Rezaee et al. [51] applied VGG-16 to recognize specific species of trees with an accuracy of 92.13% which was better than the Gradient Boosting (GB), and Random Forest (RF) algorithms that produced accuracy scores of 83.57 and 80.12%, respectively. Zhao et al. [52] demonstrated that the faster RCNN model for detection of insulator in power lines can be enhanced using anchor generation technique based on VGG-16 neural network. The enhanced performance was evident in the average precision (AP) of 0.818. Li and Guo [53] developed a VGG-16-based anomaly detection in printed circuit boards (PCB) with a best average precision (AP) score of 90% in one of its test cases. Lin and Wei [54] used a combination of the Efficient and Accurate Scene Text (EAST) algorithm and the VGG-16 deep learning network to capture information on electric pole identification plate. The proposed script recognition model recorded 83.2% accuracy. Ye et al. [55] implemented a low parameter model of VGG-16 that segmented remotely sensed images with 13% improved accuracy of the original model from 85 to 98% by reducing the input image pixel size to 64×64 pixels, and transforming the original RGB images into grayscaled images. The VGG-16 model is quite flexible to modification for better performance as can be seen in the work of Sitaula and Hossain [56]. The authors presented an attention-based VGG-16 model to detect COVID-19 in CXR images with accuracies of 79.58, 85.43, and 87.49% for three different CXR datasets, outperforming other comparable models.

In this work, the deep learning model would not be focused on a specific type of equipment, rather would provide a bird-eye-view of a substation to identify hotspots in the installation. This would be achieved by careful setting of the thermal range of the IR image acquisition system and applying the acquired images for the training of a VGG-16 deep neural network (DNN).

Methodology

VGG-16 architecture

The VGG-16 is a deep learning network developed by the Visual Geometry Group (VGG) of University of Oxford in 2014. The associated research paper by [40] came first and second in object detection and classification, respectively, of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) that same year.

As the name implies, it is made of 16 weighted layers, consisting of 13 convolutional layers and 3 fully connected layers. Both convolutional and dense layers used the ReLU activation function due to its ability to zero-out any negative neurons in the network, curtail issues of vanishing gradients, reduce amount of required computing resources,

and improve the overall efficiency of the neural network. The first layer is only for data input and has no updatable weights. It is followed by two successive convolutional layers of 64 filters each before a max pooling layer and then another two successive convolutional layers but with doubled filter size of 128. Again, a pooling layer is applied before three successive convolutional layers, this time with a filter size of 256. Another pooling layer is used as well as three convolutional layers with 512 filters.

The next three convolutional layers came after yet another pooling layer, but retains the same 512 number of filters. The last pooling layer is a 7×7 pixel matrix with 512 channels and follows two fully connected or dense layers containing one-dimensional 4096 channels. Another dense layer with one-dimensional 1000 channels is required before the last layer. The last layer of the VGG-16 network is also a one-dimensional softmax layer with 1000 outputs as specified in the ILSVRC. The one-dimensional sizes are to ensure matching between the dense and the softmax layer. The last output layer, which is the softmax layer is not counted for the network description, because it does not have trainable weights. For an RGB, the VGG-16 performs best with an input image with a tensor size of 224×224 and remains one of the most popular deep neural networks. It is represented in Fig. 1.

The filters in the VGG-16 use same padding for convolutional and maxpooling operations and are 3×3 filters, with stride of 1, also a 2×2 filters in the maxpool layer with stride of 2. The thirteen convolutional operations are performed in five stages. The first, second, third, fourth, and fifth stages have 2, 2, 3, 3, and 3 convolutional layers, respectively. The maxpooling layers are placed immediately after each convolutional block. The last three dense layers mean that there is a total of 21 layers in the VGG-16 architecture.

Hyperparameters

Hyperparameters in neural networks are the configurable parameters that determines the effectiveness and efficiency of the neural network model. They can be adjusted to enhance the model's performance. Types of hyperparameters include the batch size, number of hidden layers, number of neurons in the input and output layers. Others are the learning rate, loss function, number of iterations in each epoch, optimization technique, activation function, etc. The effect of each parameter to any model depends on the nature of the algorithm been used to develop its mathematical model, so all hyperparameters would not have same level of significant impact on a model.

In this work, the hyperparameters that played major role in the model are the number of hidden layers, loss function, activation function, batch size, optimization

Fig. 1 The standard VGG-16 deep learning model

Ukiwe *et al. Journal of Electrical Systems and Inf Technol*   (2024) 11:24

Page 8 of 25

algorithm, etc. The impact of the optimization algorithm also depends on other hyperparameters like the learning rate, as the case may be.

Activation functions

They are designed to bring in some measure of nonlinearity in the deep learning network so as to improve the aptitude of the model to comprehend complicated configurations inside the dataset, as inferred from the work of Khlongkhhoi and Kanbua [57], thereby enabling it to solve both linear and nonlinear problems. As the name implies, activation function controls the manner of activating a particular neuron in the neural network, similar to the manner of ignition of neurons in the human nervous system. The weighted sum of inputs and biases are effectively transformed to an output signal and transferred to succeeding layers in the model, thereby enabling network to decipher complicated arrangement in datasets that show nonlinear characteristics.

Epoch, batch size, and number of iteration

An epoch is the number of times required to complete a full forward and backpropagation of all the data elements through a neural network. So, in a dataset made of 500 images, setting the number of epochs to one would mean that all 500 images must be finish a full cycle of forward and backpropagation through the network. Similarly, if the number of training epoch is set as ten, then all dataset must be passed ten full cycle across network to finish the training.

Akhtar et al. [58] observed the impact of batch size in optimization algorithms. The concepts of batches and batch size, stems from the issue that most times a neural network with lots of data elements will need too much memory and computing resources to be able to process all the data during an epoch. Hence, the data are better fed to the network in specific smaller batches. The batch size is therefore the number or set of training examples in a given batch.

The number of iterations for training a network corresponds to the number of batches needed to complete one epoch. So, for a network with say n number of data examples, and k amount of batch site, then the number of iterations to complete one epoch would be given Eq. (1):

$$\text{Number of Iterations} = \text{Number of data elements } (n)/\text{Batch Size } (k) \tag{1}$$

And, the number batches reflect from the number of iterations as shown in Eq. (2):

$$\text{Number of Batches} = \text{Total Number of Iterations (in 1 epoch)} \tag{2}$$

Learning rate and optimizers

During any minimization or maximization processes, gradients are used to dictate the route of optimization of the network as it tries to converge to its optimal weights and biases by minimizing its error function. Thus, optimizers are handy for updating these weights and biases in reaction to its computed gradients.

The learning rate provides the magnitude of steps to be taken while training the neural network model. The learning determines how fast or slow the weights and biases must

be updated during the training process. Thus, the process of optimization entails seeking optimal points while traversing multi-dimensional, non-convex spaces.

In subsequent section of the work, we have evaluated the performance of the model with respect to popular and related gradient descent-based optimizers like Adam. AdaGrad, Adamax, Adadelta, and Nadam.

Adam optimizer

Adam is another name for adaptive moment estimation. It is a stochastic gradient descent (SGD) optimization technique found on an adaptive approximation of first-order and second-order moments. The method was developed by Diederik Kingma and Jimmy Ba of OpenAI and University of Toronto, respectively, in the paper presented at the International Conference on Learning and Representations 2015. In the work, the algorithm is computationally efficient, has little memory requirement, invariant to diagonal rescaling of gradients, and is well suited for problems that are large in terms of parameters.

The optimizer assesses the gradient mean as well as the squared gradient of each individual member on exponential moving average and amends any of its bias. In principle, the last weight improvement exercise is made relative to the product of the learning rate and the 1st-order moment divided by the square root of 2nd-order moment.

Hence, the algorithm uses only the learning rate, the rate of decay the gradient mean, and the rate of decomposition of the squared gradient. The authors further observed that it would be relatively easy to select the magnitude of the learning rate where the area of the best result can be predetermined, because the ultimate adjustment of the weight is roughly constrained by the learning rate. And that the exponential moving average (EMA) is skewed in the zero direction, if initialized at no value, and hence, it is divided by a definite value depending on the decomposition rate in order to arrive to a fair estimate. Adam uses the mean of the second moments of the gradients, instead of the parameter learning rates based on the average first moment as in the RMSProp.

Equation of the Adam optimizer

The Adam optimizer uses a combination of the momentum and the RMSProp. The momentum in the algorithm is applied to hasten the optimization process through the application of the exponential moving average of the gradients. The use of the mean ensures that the algorithm would converge faster toward the minima.

For the optimization of a loss function (L) with respect to its weights ω, using the traditional Stochastic Gradient Descent algorithm (SGD), whose initial weight is ω_t, its updated weight, ω_{t+1} would be computed in Eq. (3) as:

$$\omega_{t+1} = \omega_t - \alpha \left[\frac{\delta L}{\delta \omega_t} \right] \tag{3}$$

where α is the learning rate and $\partial L/\partial \omega_t$ is the rate of change of the loss function with respect to the weights w_t.

However, the single gradient can be replaced with an aggregation of multiple gradients made of exponential moving average of the previous and current gradients up to time t, called momentum, m_t.

Hence, we have SGD with momentum (SGD-M) written in Eq. (4) as follows:

$$\omega_{t+1} = \omega_t - \alpha m_t$$

(4)

where the momentum m_t is computed in Eq. (5) as:

$$m_t = \beta m_{t-1} + (1 - \beta)\left[\frac{\delta L}{\delta \omega_t}\right]$$

(5)

where β is the *moving average parameter; m_{t-1} is the previous aggregate of gradients at time t−1; and ∂L/∂ω$_t$ is the rate of change of the loss function with respect to the weights w$_t$.*

In the RMSprop-based optimizer, Eq. (6) shows that the learning rate depends on the EMA of the gradients

$$\omega_{t+1} = \omega_t - \left(\frac{\alpha}{\sqrt{v_t + \varepsilon}}\right)\left[\frac{\delta L}{\delta \omega_t}\right]$$

(6)

where ε is *a small positive constant (10^{-8})* to prevent any division by zero and V_t is the *sum of square of past gradients [i.e., sum (∂L/∂W$_{t-1}$)] at time t. It is initialized at $V_t=0$* and is calculated in Eq. (7):

$$v_t = \beta V_{t-1} + (1 - \beta) * \left[\frac{\delta L}{\delta \omega_t}\right]^2$$

(7)

And V_{t-1} represents the previous the *sum of square of past gradients [i.e., sum (∂L/∂W$_{t-1}$)] at time t−1.*

In the Adam optimizer, the EMA of the gradients (as in the case of SGD-M) is combined with the EMA of the squared gradient (as in the case of RMSprop) to obtain a different weight updating scheme of Eq. (8)

$$\omega_{t+1} = \omega_t - \left(\frac{\alpha}{\sqrt{\hat{v}_t + \varepsilon}}\right)\hat{m}_t$$

(8)

where $\hat{m}_t$ and $\hat{v}_t$ are bias-corrected values of the momentum, m_t represented in Eq. (9), and past squared gradients, v_t which is likewise computed in Eq. (10):

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}$$

(9)

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

(10)

Also, it is expressed in terms of β_1 and β_2 as per Eqs. (5) and (7), respectively, to obtain Eqs. (11) and (12)

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1)\left[\frac{\delta L}{\delta \omega_t}\right]$$

(11)

Ukiwe *et al. Journal of Electrical Systems and Inf Technol*   (2024) 11:24

Page 11 of 25

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2)\left[\frac{\delta L}{\delta \omega_t}\right]^2 \qquad (12)$$

where β_1 is the exponential decay rate for the first moment estimates (e.g., 0.9), and β_2 represents the exponential decay rate for the second-moment estimates (e.g., 0.999). These values are best set close to 1 on problems with sparse gradient (e.g., NLP and computer vision problems).

The Adam optimizer inherits and builds upon the strengths of the SGD momentum and the RMSProp to give a more optimized gradient descent. Since m_t and v_t are both initialized at 0 (based on the above methods), they have the tendency to be biased toward 0 as both β_1 and $\beta_2 \approx 1$.

The optimizer fixes this problem by computing some bias-corrected weights $\hat{m}_t$ and $\hat{v}_t$. This is also done to control the weights, while reaching the global minimum to prevent high oscillations when near it. The optimizer adapts to the gradient descent after every iteration so that it remains controlled and unbiased throughout the process, hence the name Adam.

Specific training parameters

Specifically, the algorithm calculates an exponential moving average of the gradient and the squared gradient, while the parameters β_1 and β_2 control the decay rates of these moving averages. The initial values of the moving averages β_1 and β_2 are recommended to be close to unity, resulting in a bias of moment estimates toward zero. This bias is overcome by first calculating the biased estimates before then calculating bias-corrected estimates.

(1) $\beta_1 = 0.9$
(2) $\beta_2 = 0.999$
(3) $\alpha = $ *learning rate (0.001), it was later adjusted to 0.0001 for this work*

Other settings used include 50 epochs, batch size of 16, verbose of 1, and image resized to 224×224.

Benefits of Adam optimization

One of the merits of using the Adam optimizer lies in its ease of implementation, low memory demand, and computational efficiency. Others include its ability to optimize non-stationary objects, good tendency to remain robust against diagonal rescale of objectives of the gradients, and potential to solve large-scale problems. Moreover, it is efficient for solution of sparse gradient-based optimization. Adam combines the benefits of both AdaGrad and RMSProp. RMSprop, Adadelta, and Adam are very similar algorithms that do well in similar circumstances. However, with its bias-correction factor, Adam marginally outperform RMSprop and is usually preferred.

Developing the confusion matrix

The confusion matrix provides one of the best ways of accessing the performance of a classification or detection model. Also called an error matrix, it enables good

visualization of the metrics of a model by presenting results in a square matrix in two dimensions which are the actual and the predicted classes. The number of rows and columns of the matrix depends on the number of object classes involved. And the table contains variables whereby each variable represents a combination of class and dimension. A typical confusion matrix of a binary classifier is shown in Fig. 2.

The table can be arranged with the actual predictions on the row class and the predictions on the columns and vice versa. However, they are commonly structured with the rows representing the actual values while the columns represent the predicted values. The true positive values (TP) are those predictions that the classifier correctly pointed as positive. According to Saif et al. [59], the true negative values (TN) are the ones that were correctly identified as negative by the model, whereas the false positive (FP) and false negative (FN) are the predictions that the model incorrectly labeled as positive and negative, respectively. The magnitude of the TPs, TNs, FPs, and FNs can be used independently to judge whether the model is optimized or not. For instance, high TP and TN as well as low FP and FN are important characteristics of a good model.

One of the most commonly used metrics while performing classification is accuracy. The accuracy of a classifier can be represented in Eq. (13) using values of the confusion matrix is computed according to Jia and Meng [60], as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{13}$$

The challenge of using accuracy to measure the performance of a model is that it could yield deceptive result especially on unbalanced datasets. So, we have balanced the dataset to get full benefit of the accuracy metric.

Precision and recall (sensitivity)

Other metrics like precision and recall can provide additional insight on the performance of a classification model. Precision indicates the model's ability to predict positive events by measuring how often the model achieves a correct positive prediction. It can be computed as a ratio of number of correct predictions to the number of predictions made. The recall, also called sensitivity, is applied to record the ability of the model to predict all positive outcomes, and it is also known as the sensitivity of a model. It is possible for a model to be optimized by setting the objective function to increase the recall without jeopardizing the precision. The formulas for calculating precision and recall (sensitivity) are also given by Hama and Omer [61] in Eqs. (14) and (15):

		PREDICTIONS	
		+VE	-VE
ACTUALS	+VE	TP	FN
	-VE	FP	TN

Fig. 2 Confusion matrix of a typical binary classifier

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{14}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{15}$$

Another important metrics for analyzing performance of classification model is the F_β measures. The F_β-score is computed with the recall and precision scores of the model's confusion matrix, as presented by Tadist et al. [62]. It is achieved by applying the weighted average of the recall and the precision, as shown in equation (16):

$$F_\beta = (1 + \beta^2) \times \frac{(\text{precision} \times \text{recall})}{(\beta^2 \times \text{precision}) + \text{recall}} \tag{16}$$

When the magnitude of β is set as 1, the result is referred to as the F_1 measure and calculated as in Eq. (17):

$$F_1 = 2 \times \frac{(\text{precision} \times \text{recall})}{(\text{precision} + \text{recall})} \tag{17}$$

Global average pooling (GAP)

This is a type of pooling function made to replace fully connected or dense layers in a typical deep learning model. The essence is to generate one feature map for each corresponding category of the classification task at the last multilayer perceptron (MLP) like layer. So rather than introducing dense layers after the feature maps, the mean of a specific feature map is selected, and the resulting vector serves as the input to the softmax layer. In the breakthrough work of Lin et al. [63], the author noted that merit of the GAP operation against the dense layers lies in the fact that it is similar and more amenable to the convolutional structure of the preceding layers, thereby streamlining interactions between the feature maps and the output categories. Another benefit of the GAP layer is that overfitting is evaded with this layer in place because no parameter can be optimized therein. Moreover, the GAP operation is more vigorous to spatial translations of the input because it totals out spatial information of the feature map.

Unlike Flattening operation that transforms a tensor of any dimension into a one-dimensional array, and retaining each value of the array, GlobalAveragePooling2D (GAP-2D) executes average pooling on the spatial dimensions until each spatial dimension is one, and leaves other dimensions unchanged. The GAP is a technique for superior depiction of the vector. It can be applied in different dimensions and uses an unweighted kernel window to slide over each feature map and pools the pixel values therein, by computing the mean. Padding is done to ensure pixels at the edges of the feature maps are not left-out.
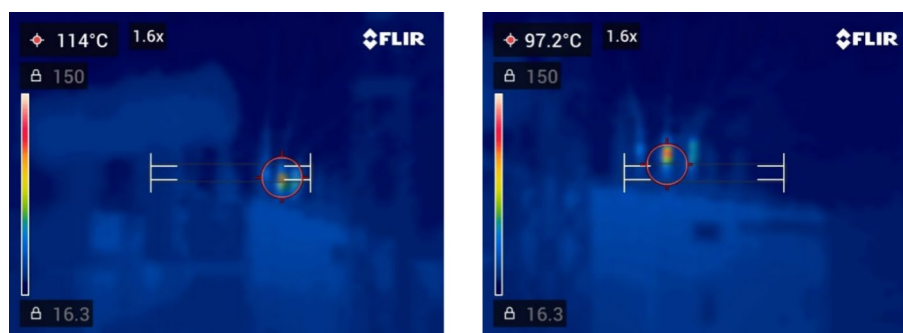
Infrared image acquisition

Infrared images of live electrical equipment installed in Nigeria's various 132/33 kV transmission substations were captured using the FLIR C5 compact infrared thermal camera. Typical images acquired with the device measures 640×480 pixel. The upper

thermal measuring range was set at 150 °C, while the lower thermal limit was kept at 16.3 °C. The rainbow color palette was selected. According to FLIR Teledyne [64], renowned for manufacturing IR imaging systems, this color scheme begins with deep blue hues for the lower pixel values and progresses to the higher pixel values, indicated by the transition from dark to the brighter colors. It has more colors in its mix and offers good background for highlighting points therein with little thermal differential. Thermal images of power equipment were taken at various distances within the specified range of the IR camera. Since, each image were taken under same standard camera setting, the color bars accompanying them and the proprietary FLIR logo can be removed without altering the training result or interpretation by the computer vision-based deep learning model.

IR images of different rating of the same type of equipment were also obtained to improve the performance of the deep learning network. Moreover, the IR images were acquired at different weather conditions in order to counter issues of reflected temperature from the surfaces of the equipment. This is possible because the IR radiation is not affected by the amount of illumination around the object under observation. The FLIR C5 infrared imager comes with a zoom capability. This means that hotspots can be captured from safe distance by using the zoom function, thereby ensuring the safety of the operator. Figure 3 shows sample of images captured under the stated thermal camera settings, which highlights the occurrence of electrical hotspots of 114 and 97.2 °C on the bushings of a 15 MVA, 33/11 kV transformer with a $1.6 \times$ zoom function.

The equipment affords the opportunity to lock the temperature range within a specific upper and lower limit during the image acquisition process as shown in Fig. 3. The padlock icon beside the color bar shows that the temperature range were restricted from changing during the process of image acquisition. The IR imaging device can last up to 4 h on full battery charge based on the manufacturer's specification, but due to the length of time required for seamless data acquisition, additional 2 nos. 20,000 mah Lithium powered power banks were used as backup for the built-in power supply.

The images were acquired at different times from diverse power substations with various configurations within the national grid. The acquired raw IR image datasets were further processed by labeling, cleaning, grouping and augmenting the data within each group. The IR image augmentation were done using Keras-based Tensorflow libraries using python packages in Jupyter Notebook, to balance and enhance the amount of data

Fig. 3 Sample IR image with set thermal limits

Ukiwe *et al. Journal of Electrical Systems and Inf Technol*   (2024) 11:24

Page 15 of 25

available for the deep learning network. The process flow for implementation of the deep learning model is shown in Fig. 4.

The final dataset was made of 1000 IR images of power equipment in a transmission substation, evenly divided into 500 normal images without hotspots and 500 images with hotspots. And the test dataset of 100 images was used to test the model for hotspot identification by drawing a box around the region of interest containing the hotspot. The model was executed using different optimizers, with an instruction to identify (or not) up to five hotspots in an image and the results compared. The merit of this approach is the possibility of the human visual sense having difficulty in identifying the least manifestation of hotspot, which the computer vision-based application could easily achieve.

Results and discussion

The results of the hotspot identification using the VGG-16 deep learning network indicates the ability of the global average pooling layer to segregate high pixel values out of the neighboring ones, thereby identifying them as an anomaly within the IR image, and taken those areas as the region(s) of interest. Different variants of the Adam optimizers were used to implement the model and the results were compared.
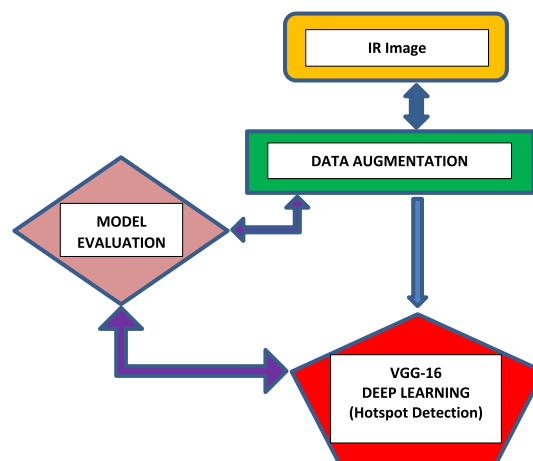
Results with Adamax optimizer
The Adamax optimizer yielded a good result as shown in Fig. 5.
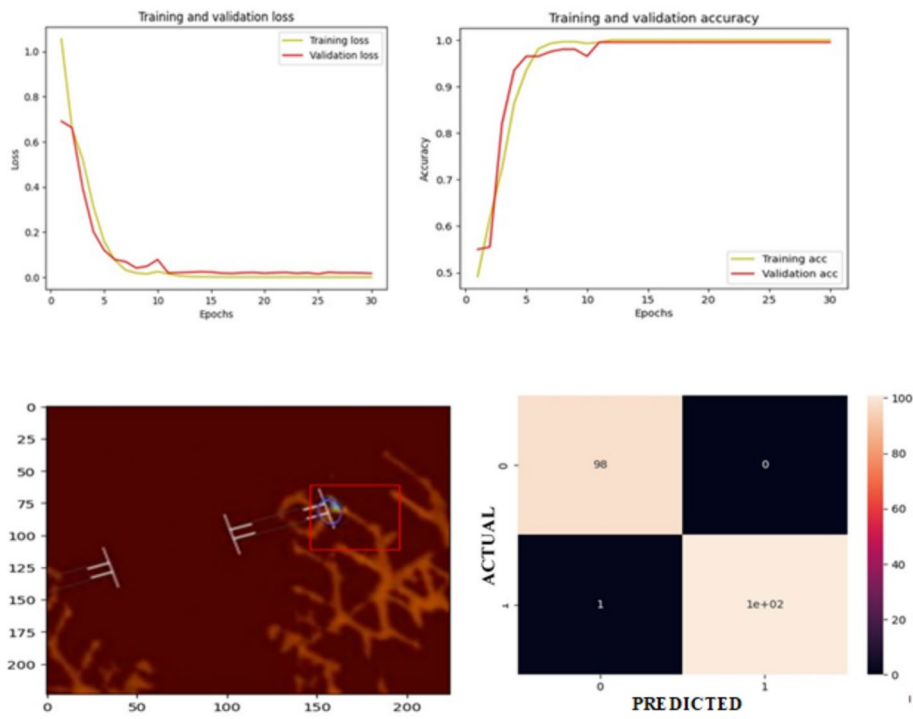
Results with Adadelta optimizer
The Adadelta optimizer proposed by Zeiler [65] was used to execute the model, and the result is shown in Fig. 6.
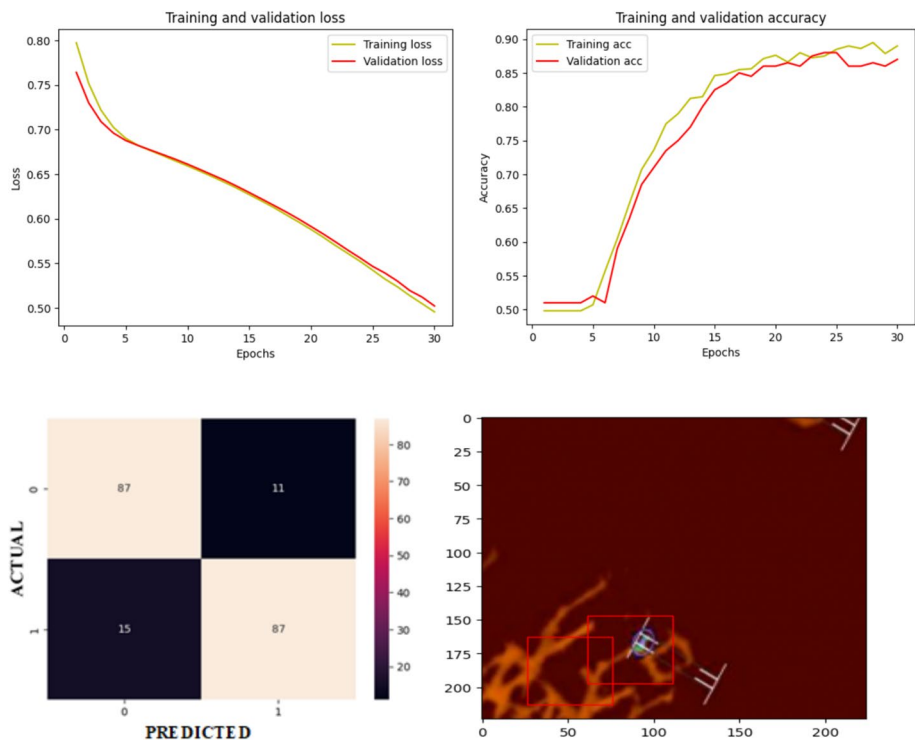
Results with AdaGrad optimizer
The AdaGrad optimizer as proposed by Duchi et al. [66] was used to execute the VGG-16 hotspot detection model. The result obtained is shown in Fig. 7. The model was instructed to identify up to five hotspots from randomly selected test IR images, and the result can be seen from Figs. 6 and 7.

Fig. 4 Process flow diagram of the hotspot detection research using VGG-16 deep neural network
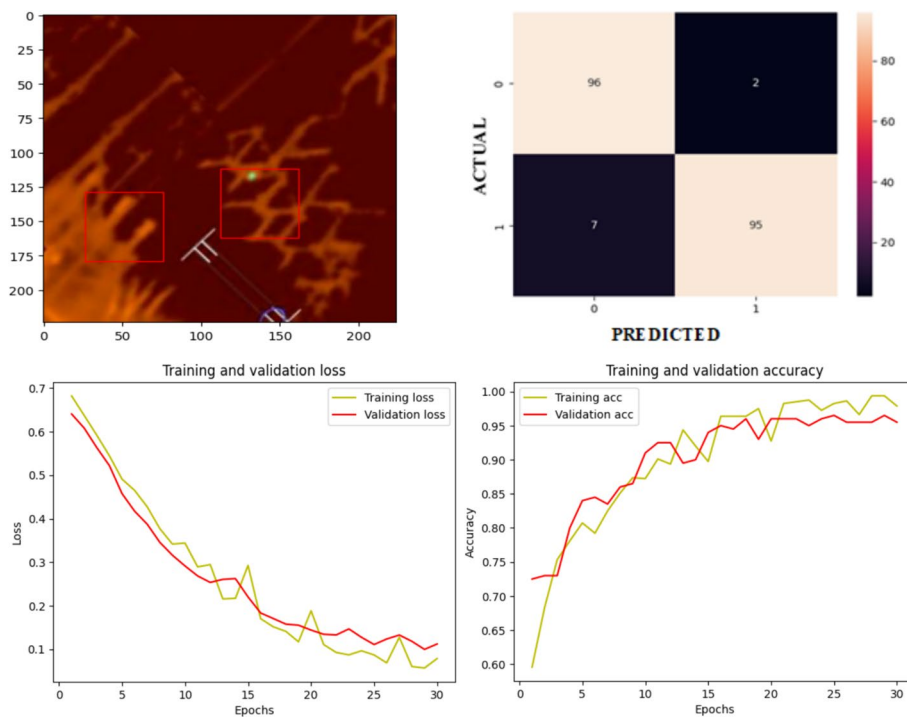
Fig. 5 Results of VGG-16 hotspot detection with Adamax optimizer

Fig. 6 Results of hotspot detection in power installation with Adadelta

Fig. 7 Results of VGG-16 implementation of hotspot detection with AdaGrad

Results with Nadam optimizer

The Adam algorithm was combined with the Nesterov Momentum by Dozat [67] to yield a new optimizer called Nesterov-accelerated adaptive moment (Nadam). The optimizer was also used to test the hotspot detection model with good result shown in Fig. 8.

**Results of Adam optimizer implementation with learning rate of 0.001**

The model performance was not as good at learning rate of 0.001 proposed in the work of Kingma et al. [68] as shown in Fig. 9 at 51% accuracy. This validates the fact that some of the hyperparameters applied in some breakthrough research may not be exactly applicable to future works. There may be need for tuning of these hyperparameters for optimal result in peculiar works.

**Results with Adam optimizer with learning rate of 0.0001**

The model recorded spectacular performance of 99.98% accuracy when implemented with Adam using a learning rate of 0.0001 as presented in Fig. 10. This is because as per the literature, reducing the learning rate improves the pixel-to-pixel discriminating ability of the model.

**Comparison with related research**

The comparative result of the work presented in Table 1 shows that the Adam optimizer yielded the best performance of 99.98% accuracy with learning rate of 0.0001, better than other similar optimizers for the hotspot detection model.

Ukiwe *et al. Journal of Electrical Systems and Inf Technol* (2024) 11:24

Page 18 of 25
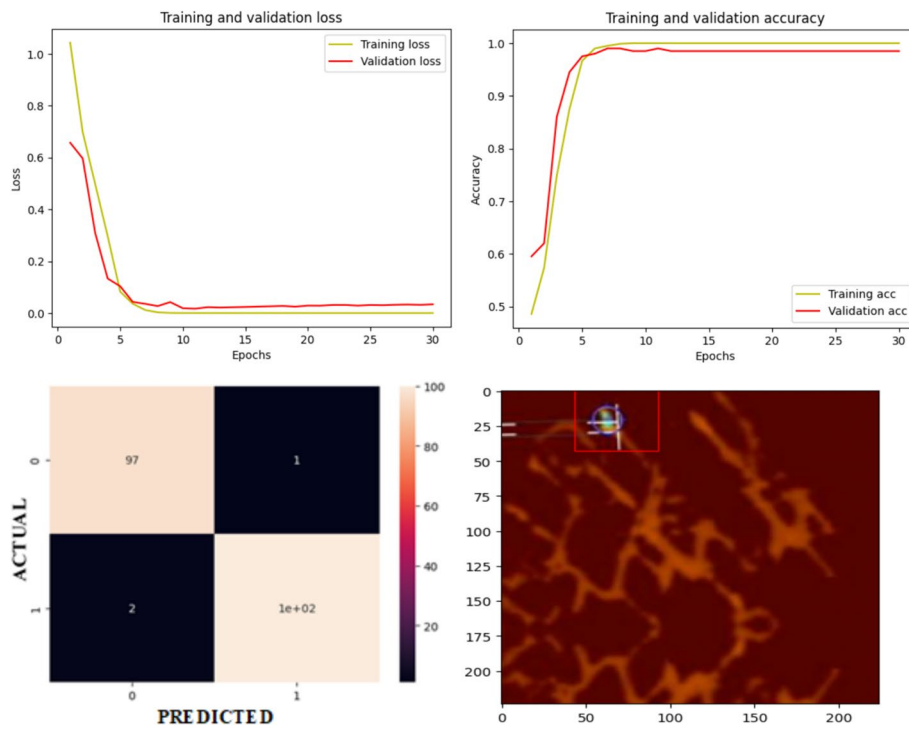
Fig. 8 Results of VGG-16 implementation of hotspot detection with Nadam
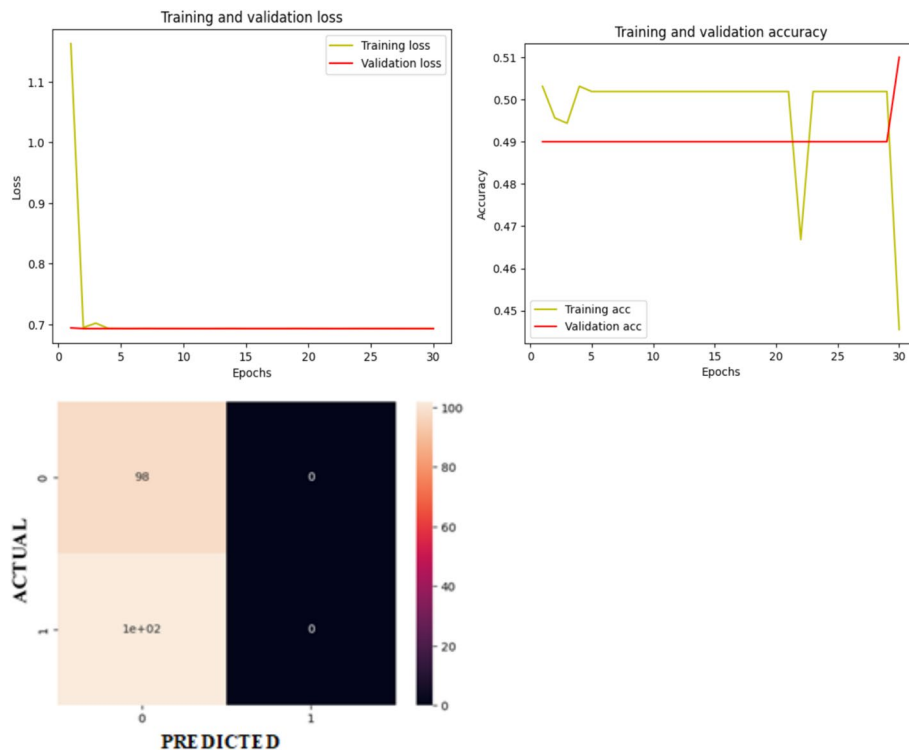
Fig. 9 Adam implementation of hotspot detection model with 0.001 learning rate
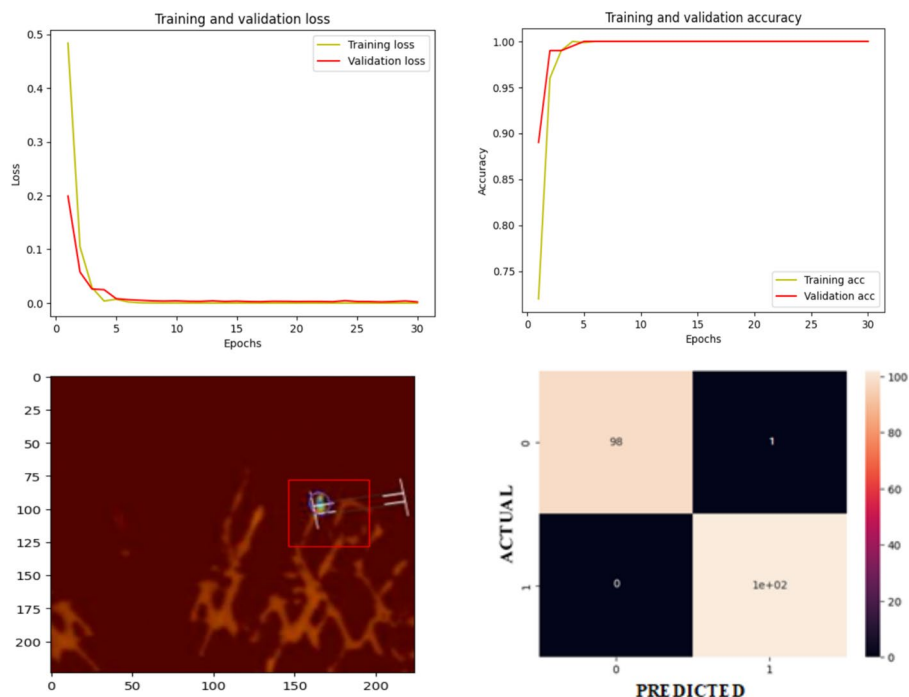
Fig. 10 Adam implementation of VGG-16 hotspot detection model with 0.0001 learning rate

Table 1 Comparative analysis of VGG-16 model for hotspot detection using different optimizers

Metrics	Comparative analysis of VGG-16 model for hotspot detection using different optimizers				
	Adamax	Adadelta	AdaGrad	Nadam	Adam
Accuracy %	99.50	87.00	95.49	98.50	99.98
Precision	0.9898	0.8529	0.9320	0.9798	1.0000
Recall	1.0000	0.8876	0.9796	0.9898	0.9898
F1-Score	0.9945	0.8699	0.9552	0.9798	0.9949

A comparison of results obtained from similar previous research is shown in Table 2.

Table 2 indicates that the work by Kayci et al. [69] utilized YOLOv3 with 97% accuracy but limited the work to only solar PV modules. But Zheng et al. [73] adopted the S-YOLOv5 method of hotspot detection also in solar PV modules with mean average precision (mAP) of 98.1%. However, Venkatesh and Sugumaran [75] used visual images only on VGG-16 model to segregate good and bad panels, whereas Pierdicca et al. [76] applied the VGG-16 model for isolating faults in solar PV panels were geared toward classifying the panels into normal and damaged without identifying the region plagued by hotspot. Some of the merits in our implemented pre-trained VGG-16 DL-based hotspot detection model using GAP layer in place of the fully connected layer besides outperforming other similar models with a test accuracy of 99.98%, resides in the prospect of actually identifying multiple hotspots in the system,

Ukiwe *et al. Journal of Electrical Systems and Inf Technol*   (2024) 11:24

Page 20 of 25

Table 2 Comparison of results from previous related works

Title/Reference	Doi	Method/Results
Deep learning-based fault detection and diagnosis in photovoltaic system using thermal images acquired by UAV [69]	https://doi.org/10.2339/politeknik.1094586	The target equipment was Solar PV. YOLOv3 was applied with 97% accuracy detection of cell fault, 95% in detection of module fault, and 93% in detection of panel fault
A Comparative Study of MLP Networks Using Backpropagation Algorithms in Electrical Equipment Thermography [70]	https://doi.org/10.1007/s13369-014-0989-7	IR images of power distribution equipment were obtained with Fluke Ti25. MLP with backpropagation after training with RP showed best training and testing accuracy of 84.21% and 76.20%, respectively. Best training sensitivity 80% and testing sensitivity of 67.50%. Also, the lowest FN of 20% and 32.50% for training and testing phases, respectively
Photovoltaic Hot-Spots Fault Detection Algorithm using Fuzzy Systems [71]	https://doi.org/10.1109/TDMR.2019.2944793	Dataset of Solar PV came from Solar UK Database. FIS using Mamdani-type fuzzy logic controller was modeled with three inputs: open circuit voltage, short circuit current, and percentage of power loss. Tested average accuracy of 96.7%.was obtained
Fault detection using thermal image based on soft computing methods: Comparative study [72]	https://doi.org/10.1016/j.microrel.2017.02.013	FLIR SC4000 was used to get the dataset of integrated circuits. Study focused on fault detection in PCB with FEM. Extracted effective histogram features of the ICs hotspots were minimized using PCA trained a classifier model. Results were compared with MLP, SVM, and ANFIS models. MLP (train 96.32 test 88.90); ANFIS (train 97.44 test 94.18); SVM (train 88.54 test 97.43)
Lightweight Hot-Spot Fault Detection Model of Photovoltaic Panels in UAV Remote-Sensing Image [73]	https://doi.org/10.3390/s22124617	Study focused on Solar PV. ShuffleNetv2-YOLOv5 (S-YOLOv5) was used, and the original feature fusion method was abridged with mAP of 98.1%
Novel Photovoltaic Hotspotting Fault Detection Algorithm [74]	https://doi.org/10.1109/TDMR.2019.2910196	Hot-spots in Solar PV systems were isolated using CDF modeling, the CDF model predicted presence of PV hot-spots in the PV modules. The CDF models got the PV hotspots with 80% accuracy
Fault Detection in aerial images of photovoltaic modules based on Deep learning [75]	https://doi.org/10.1088/1757-899X/1012/1/012030	Proposed a fault detection in Solar PV module using DL. The feature extraction and fault classification was implemented using VGG-16 pre-trained network resulting to mean training accuracy of 98.47% and validation accuracy of 95.40%
Deep Convolutional Neural Network for Automatic Detection of Damages Photovoltaic Cells [76]	https://doi.org/10.5194/isprs-archives-XLII-2-893-2018	Model was based on a Balanced Dataset of Solar PV modules acquired using FLIR TAU 2; predicted damaged module with Precision of 0.83, Recall of 0.50, and F1-score of 0.63. Normal modules were predicted with a Precision of 0.64, Recall of 0.90, and F1-score of 0.75

because the model can be configured to identify more than one region containing hotspots in an IR image of a power system.

Conclusion and future work

The proposed transfer learning-based VGG-16 deep learning model achieved an accuracy measure of up to 99.98% using the Adam optimizer. Careful choice of the 16.3–150 °C thermal range together with the rainbow color palette made for standardized IR training images that helps the models to easily pinpoint the hotspots within the electrical power system. Another spectacular feature of the model is that it can figure out very tiny hotspots that may not necessarily be visible using the human visual system, which is one of the essence of computer vision.

Applicability and limitations

The presented approach converges within finite time, with low computational complexity. Hence, it is implementable with moderate computing infrastructure and would reduce the dependency on power system experts for timely identification of hotspots in electrical installations.

It can be applied for multiple hotspot detection in most types of electrical/electronic devices or installations like rotating machines, solar power equipment, power electronics, etc. In leaking petrochemical fluid or gas storage systems, anomalies in the storage vessel can be evident in the form of temperature differential around the compromised area. Moreover, in monitoring heating, ventilating, and air conditioning (HVAC) or power generating systems with feedstock its prospects are quite enormous because temperature difference would usually exist between area containing fluids in pipes or containers and regions containing gases. Such thermal profile can be used to segregate anomalies. In roofing systems, the method can be used to isolate tiny holes prone to leakages. In fact, the applications in thermal anomaly detection cannot be exhausted.

However, limitations exist in that the model depends on training with carefully taken IR images using cautiously selected thermal range to make it easy for the model to pinpoint any anomaly therein.

Future research areas

The work has many prospects for future research. It can be implemented using various deep learning models like ResNet, AlexNet, EfficientNet, etc. Moreover, there is potential for exploring implementation of these models using the GlobalMaxPooling in place of the GlobalAveragePooling Layer.

Abbreviations

Adadelta	Adaptive delta
AdaGrad	Adaptive gradient
Adam	Adaptive moment estimation
Adamax	Adam-Max
AI	Artificial intelligence
ANFIS	Adaptive neuro-fuzzy inference system
ANN	Artificial neural network
AP	Average precision
AUROC	Area under the ROC curve
BFG	Broyden–Fletcher–Goldfarb–Shanno quasi-Newton
CDF	Cumulative density function

CNN	Convolutional neural network
CXR	Chest X-ray
DCNN	Deep convolutional neural network
DL	Deep learning
DNN	Deep neural network
EAST	Efficient and Accurate Scene Test
EMA	Exponential moving average
Faster RCNN	Faster region-based CNN
FEM	Finite element modeling
FIS	Fuzzy inference system
FN	False negative
FP	False positive
GAP	Global average pooling
GAN	Generative adversarial network
GB	Gradient boosting
GD	Gradient descent
GDX	Gradient descent with momentum and adaptive learning rate
GFCC	Gammatone frequency cepstral coefficient
IC	Integrated circuit
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
IR	Infrared
IRT	Infrared thermography
LM	Levenberg–Marquardt
mAP	Mean average precision
Mask RCNN	Mask region CNN
MLP	Multilayer perceptron
MOSA	Metal oxide surge arrester
MRI	Magnetic resonance imaging/image
Nadam	Nesterov-accelerated adaptive moment
PCA	Principal component analysis
PCB	Printed circuit board
PCNN	Pulse coupled neural network
PV	Photovoltaic system
RGB	Red green blue
ROC Curve	Receiver operating characteristic curve
ROI	Region of interest
ReLU	Rectified linear unit
ResNet	Residual network
RF	Random forest
RMSprop	Root mean square propagation
RP	Resilient backpropagation
SCADA	Supervisory control and data acquisition
SGD	Stochastic gradient descent
SGD (M)	Stochastic gradient descent with momentum
TN	True negative
TP	True positive
VA-GAN	Variational autoencoder GAN
VGG	Visual geometry group
WAR DIC	Wasserstein autoencoder reconstruction-differential image classifier
YOLO	You only look once
ZnO	Zinc oxide

Availability of data and materials
The data that support the findings of this study are available from the corresponding author, upon reasonable request. Restrictions apply to the availability of these data, which were used with permission of the Transmission Company of Nigeria for this study.

Declarations

Competing interests
All authors certify that they have no affiliations with or involvement in any organization or entity with any financial or non-financial interest in the subject matter or materials discussed in this manuscript.

References

1. Adedigba AP, Adeshina SA, Aibinu AM (2022) Performance evaluation of deep learning models on mammogram classification using small dataset. Bioengineering 9:161. https://doi.org/10.3390/bioengineering9040161
2. Oluseyi P, Adeagbo J, Dinakin DD, Akinbulire TO (2020) Mitigation of hotspots in electrical components and equipment using an adaptive neuro-fuzzy inference system. Electr Eng 102:8. https://doi.org/10.1007/s00202-020-01028-0
3. Sousa E, Vardasca R, Teixeira S, Seixas A, Mendes J, Costa-Ferreira A (2017) A review on the application of medical infrared thermal imaging in hands. Infrared Phys Technol 85:315–323. https://doi.org/10.1016/j.infrared.2017.07.020
4. Usamentiaga R, Pablo V, Guerediaga J, Vega L, Molleda J, Bulnes FG (2014) Infrared thermography for temperature measurement and non-destructive testing. Sensors 14(7):12305–12348. https://doi.org/10.3390/s140712305
5. Aidossov N, Zarikas V, Zhao Y, Mashekova A, Ng EY, Mukhmetov O, Mirasbekov Y, Omirbayev A (2023) An integrated intelligent system for breast cancer detection at early stages using IR images and machine learning methods with explainability. SN Comput Sci 4:184. https://doi.org/10.1007/s42979-022-01536-9
6. Ng EYK (2009) A review of thermography as promising non-invasive detection modality for breast tumor. Int J Therm Sci 48(5):849–859. https://doi.org/10.1016/j.ijthermalsci.2008.06.015
7. Epperly RA, Heberlin GE, Eads LG (1999) Thermography, a tool for reliability and safety. IEEE Ind Appl Mag 5(1):28–36. https://doi.org/10.1109/2943.740757
8. Balakrishnan GK, Yaw CT, Koh SP, Abedin T, Raj AA, Tiong SK, Chen CP (2022) A review of infrared thermography for condition-based monitoring in electrical energy: applications and recommendations. Energies 15:6000. https://doi.org/10.3390/en15166000
9. Ukiwe EK, Adeshina SA, Tsado J (2023) Techniques of infrared thermography for condition monitoring of electrical power equipment. J Electr Syst Inf Technol 10:49. https://doi.org/10.1186/s43067-023-00115-z
10. Gao Z, Zhang Y, Li Y (2020) Extracting features from infrared images using convolutional neural networks and transfer learning. Infrared Phys Technol 105:103237. https://doi.org/10.1016/j.infrared.2020.103237
11. Choi RY, Coyner AS, Cramer JK, Chiang MF, Campbell JP (2020) Introduction to machine learning, neural networks, and deep learning. Transl Vis Sci Technol 9:2. https://doi.org/10.1167/tvst.9.2.14
12. Ahmed A, Rio-Chanona EAD, Mercangoz M (2022) Learning linear representations of nonlinear dynamics using deep learning. Int Fed Autom Control IFAC PapersOnLine 55:162–169. https://doi.org/10.1016/j.ifacol.2022.07.305
13. Wang X, Zhao Y, Pourpanah F (2020) Recent advances in deep learning. Int J Mach Learn Cybern 11:747–750. https://doi.org/10.1007/s13042-020-01096-5
14. Balla A, Habaebi MH, Islam MR, Mubarak S (2022) Applications of deep learning algorithms for supervisory control and data acquisition intrusion detection system. Clean Eng Technol 9:100532. https://doi.org/10.1016/j.clet.2022.100532
15. Soni R (2016) Computer vision. Int J Comput Sci Inf Technol Res 4:76–81
16. Wiley V, Lucas T (2018) Computer vision and image processing: a paper review. Int J Artif Intell Res 2(1):28–36. https://doi.org/10.29099/ijair.v2i1.42
17. Rybchak Z, Basystiuk O (2017) Analysis of computer vision and image analysis technics. ECONTECHMOD Int Q J 6(2):79–84
18. Acquaah YT, Gokaraju B, Tesiero RC III, Monty GH (2021) Thermal imagery feature extraction techniques and the effects on machine learning models for smart HVAC efficiency in building energy. Remote Sens 13:3847. https://doi.org/10.3390/rs13193847
19. Mustapha A, Mohamed L, Ali K (2021) Comparative Study of optimization techniques in deep learning: application in the ophthalmology field. In: The international conference on mathematics & data science (ICMDS), Journal of physics: conference series, vol 1743, p 012002. https://doi.org/10.1088/1742-6596/1743/1/012002
20. Adedigba AP, Adeshina SA, Aina OE, Aibinu AM (2021) Optimal hyperparameter selection of deep learning models for COVID-19 chest X-ray classification. Intell Based Med 5:100034. https://doi.org/10.1016/j.ibmed.2021.100034
21. Yang J, Wang W, Lin G, Sun QL, Sun Y (2019) Infrared thermal imaging-based crack detection using deep learning. IEEE Access 7:182060–182077. https://doi.org/10.1109/ACCESS.2019.2958264
22. Ding W, Xia Y, Wang Z, Chen Z, Gao X (2020) An ensemble-learning method for potential traffic hotspots detection on heterogeneous spatio-temporal data in highway domain. J Cloud Comput Adv Syst Appl 9:25. https://doi.org/10.1186/s13677-020-00170-1
23. Fang Q, Castanedo CI, Garrido I, Duan Y, Maldague X (2023) Automatic detection and identification of defects by deep learning algorithms from pulsed thermography data. Sensors 23:4444. https://doi.org/10.3390/s23094444
24. Choudhary A, Mian T, Fatima S (2021) Convolutional neural network based bearing fault diagnosis of rotating machine using thermal images. Measurement 176:109196. https://doi.org/10.1016/j.measurement.2021.109196
25. Das AK, Dey D, Chatterjee B, Dalai S (2021) A transfer learning approach to sense the degree of surface pollution for metal oxide surge arrester employing infrared thermal imaging. IEEE Sens J 21:16961–16968. https://doi.org/10.1109/JSEN.2021.3079570

Ukiwe *et al. Journal of Electrical Systems and Inf Technol*   (2024) 11:24

Page 24 of 25

26. Chandra S, AlMansoor K, Chen C, Shi Y, Hyungjoon (2022) Deep learning based infrared thermal image analysis of complex pavement defect conditions considering seasonal effect. Sensors 22:9365. https://doi.org/10.3390/s22239365

27. Janssens O, Loccufier M, Van de Walle R, Van Hoecke S (2018) Deep learning for infrared thermal image based machine health monitoring. IEEE ASME Trans Mechatron 23(1):151–159. https://doi.org/10.1109/TMECH.2017.2722479

28. Fanchiang KH, Huang YC, Kuo CC (2021) Power electric transformer fault diagnosis based on infrared thermal images using Wasserstein generative adversarial networks and deep learning classifier. Electronics 10:1161. https://doi.org/10.3390/electronics10101161

29. Jiang J, Bie Y, Li J, Yang X, Ma G, Lu Y, Zhang C (2021) Fault diagnosis of the bushing infrared images based on mask R-CNN and improved PCNN joint Algorithm. High Volt 6:116–124. https://doi.org/10.1049/hve.2019.0249

30. Fanchiang KH, Kuo CC (2022) Application of thermography and adversarial reconstruction anomaly detection in power cast-resin transformer. Sensors 22:1565. https://doi.org/10.3390/s22041565

31. Fang J, Yang F, Tong R, Yu Q, Dai X (2021) Fault diagnosis of electric transformers based on infrared image processing and semi-supervised learning. Sci Dir Glob Energy Interconnect 4(6):596–607. https://doi.org/10.1016/j.gloei.2022.01.008

32. Mlakić D, Nikolovski S, Majdandžić L (2018) Deep learning method and infrared imaging as a tool for transformer faults detection. J Electr Eng 6:98–106. https://doi.org/10.17265/2328-2223/2018.02.006

33. Jangblad M (2018) Object detection in infrared images using deep convolutional neural network. Uppsala University, UPTEC F 18028, Examensarbete 30 hp. https://www.diva-portal.org/smash/get/diva2:1228617/FULLTEXT01.pdf

34. Jaffery ZA, Dubey AK (2014) Design of early fault detection technique for electrical assets using infrared thermograms. Electr Power Energy Syst 63:753–759

35. Kumar D, Ansari MA (2018) Condition monitoring of electrical assets using digital IRT and AI technique. J Electr Syst Inf Technol 5:623–634. https://doi.org/10.1016/j.jesit.2017.10.001

36. Novizona B, Maleka ZA, Bashira N, Asilaha N (2013) Thermal image and leakage current diagnostic as a tool for testing and condition monitoring of ZnO surge arrester. J Teknol Sci Eng 64(4):27–32

37. Alvarado-Hernandez AI, Zamudio-Ramirez I, Jaen-Cuellar AY, Osornio-Rios RA, Donderis-Quiles V, Antonino-Daviu JA (2022) Infrared thermography smart sensor for the condition monitoring of gearbox and bearings faults in induction motors. Sensors 22:6075. https://doi.org/10.3390/s22166075

38. Parashar S, Kumar A, Sharma P, Rana S, Kumar D (2023) Fault prediction in electrical assets using infrared thermography. In: Advancements & Key challenges in green energy and computing (AKGEC 2023), Journal of physics: conference series, vol 2570, p 012019. https://doi.org/10.1088/1742-6596/2570/1/012019

39. Liu Z, Wang J, Duan L, Shi T, Fu Q (2017) Infrared image combined with CNN based fault diagnosis for rotating machinery. In: 2017 international conference on sensing, diagnostics, prognostics, and control, pp 137–142. https://doi.org/10.1109/SDPC.2017.35

40. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: Conference paper at international conference on learning representation. https://doi.org/10.48550/arXiv.1409.1556

41. Jiang L, Yi W (2022) Power grid fault diagnosis method based on VGG network line graph semantic extraction. Int J Sci Eng Res IJSER 10:16–20

42. Younis A, Qiang L, Nyatega CN, Adamu MJ, Kawuwa HB (2022) Brain tumor analysis using deep learning and VGG-16 ensembling learning approaches. Appl Sci 12:7282. https://doi.org/10.3390/app12147282

43. Piekarski M, Korjakowska JJ, Wawrzyniak AI, Gorgon M (2020) Convolutional neural network architecture for beam instabilities identification in synchrotron radiation systems as an anomaly detection problem. Measurement 165:108116. https://doi.org/10.1016/j.measurement.2020.108116

44. Dang XJ, Wang FH, Ma WJ (2002) Fault diagnosis of power transformer by acoustic signals with deep learning. In: 2020 IEEE international conference on high voltage engineering and application (ICHVE), Beijing, China, pp 1–4. https://doi.org/10.1109/ICHVE49031.2020.9279751.

45. Sharma S, Guleria K (2023) A deep learning based model for the detection of pneumonia from chest X-ray images using VGG-16 and neural networks. In: International conference on machine learning and data engineering, procedia computer science, vol 218, pp 357–366. https://doi.org/10.1016/j.procs.2023.01.018

46. Du B, He Y, Duan J, Zhang Y (2020) Intelligent classification of silicon photovoltaic cell defects based on eddy current thermography and convolution neural network. IEEE Trans Ind Inf 16(10):6242–6251. https://doi.org/10.1109/TII.2019.2952261

47. Alatawi AA, Alomani SM, Alhawiti NI, Ayaz M (2022) Plant disease detection using AI based VGG-16 model. Int J Adv Comput Sci Appl IJACSA 13(4):718–727

48. Ahmad I, Hamid M, Yousaf S, Shah ST, Ahmad MO (2020) Optimizing pretrained convolutional neural networks for tomato leaf disease detection. Hindawi Complex 2020:8812019. https://doi.org/10.1155/2020/8812019

49. Rasyid I, Yudianto MRA, Maimunah, Purnomo TA (2023) Electronic product recommendation system using the cosine similarity algorithm and VGG-16. Sink J Penelit Tek Inf 8(4):2120–2129. https://doi.org/10.33395/sinkron.v8i4.12936

50. Sheriff STM, Kumar JV, Vigneshwaran S, Jones A, Anand J (2021) Lung cancer detection using VGG NET 16 architecture. In: International conference on physics and energy 2021 (ICPAE 2021), Journal of physics: conference series, p 012001. https://doi.org/10.1088/1742-6596/2040/1/012001

51. Rezaee M, Zhang Y, Mishra R, Tong F, Tong H (2018) Using a VGG-16 network for individual tree species detection with an object-based approach. In: 2018 10th IAPR workshop on pattern recognition in remote sensing (PRRS), Beijing, China, pp 1–7. https://doi.org/10.1109/PRRS.2018.8486395.

52. Zhao Z, Zhen Z, Zhang L, Qi Y, Kong Y, Zhang K (2019) Insulator detection method in inspection image based on improved faster R-CNN. Energies 12:1204. https://doi.org/10.3390/en12071204

53. Li YT, Guo JI (2018) A VGG-16 based faster RCNN model for PCB error inspection in industrial AOI applications. In: 2018 IEEE international conference on consumer electronics-Taiwan (ICCE-TW), Taichung, Taiwan, pp 1–2. https://doi.org/10.1109/ICCE-China.2018.8448674

Ukiwe *et al. Journal of Electrical Systems and Inf Technol*   (2024) 11:24

Page 25 of 25

54. Lin S, Wei Q (2020) Study on text detection and positioning method of utility pole identification plate based on improved EAST. In: 2020 IEEE 4th information technology, networking, electronic and automation control conference (ITNEC), Chongqing, China, pp 2374–2379. https://doi.org/10.1109/ITNEC48623.2020.9084779

55. Ye M, Ruiwen N, Chang Z, He G, Tianli H, Shijun L, Yu S, Tong Z, Ying G (2021) A lightweight model of VGG-16 for remote sensing image classification. IEEE J Sel Top Appl Earth Obs Remote Sens 14:6916–6922. https://doi.org/10.1109/JSTARS.2021.3090085

56. Sitaula C, Hossain MB (2021) Attention-based VGG-16 model for COVID-19 chest X-ray image classification. Appl Intell 51:2850–2863. https://doi.org/10.1007/s10489-020-02055-x

57. Khlongkhoi P, Chayantrakom K, Kanbua W (2019) Application of a deep learning technique to the problem of oil spreading in the Gulf of Thailand. Adv Differ Equ 2019:306. https://doi.org/10.1186/s13662-019-2241-y

58. Akhtar MU, Raza MH, Shafiq M (2019) Role of batch size in scheduling optimization of flexible manufacturing system using genetic algorithm. J Ind Eng Int 15:135–146. https://doi.org/10.1007/s40092-018-0278-2

59. Saif D, Sarhan AM, Elshennawy NM (2024) Early prediction of chronic kidney disease based on ensemble of deep learning models and optimizers. J Elect Syst Inf Technol 11:17. https://doi.org/10.1186/s43067-024-00142-4

60. Jia X, Meng MQH (2016) A deep convolutional neural network for bleeding detection in wireless capsule endoscopy images. In: 38th annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA, pp 639–642. https://doi.org/10.1109/EMBC.2016.7590783

61. Hama HM, Abdulsamad TS, Omer SM (2024) Houseplant leaf classification system based on deep learning algorithms. J Electr Syst Inf Technol 11:18. https://doi.org/10.1186/s43067-024-00141-5

62. Tadist K, Mrabti F, Nikolov NS et al (2021) SDPSO: spark distributed PSO-based approach for feature selection and cancer disease prognosis. J Big Data 8:19. https://doi.org/10.1186/s40537-021-00409-x

63. Lin M, Chen Q, Yan S (2012) Network in network. arXiv preprint arXiv:1312.4400

64. FLIR TELEDYNE (2021) Picking a Color Palette. https://www.flir.com/discover/industrial/picking-a-thermal-color-palette/

65. Zeiler MD (2012) Adadelta: an adaptive learning rate method. https://arxiv.org/abs/1212.5701

66. Duchi J, Hazan E, Singer Y (2011) Adaptive subgradient methods for online learning and stochastic optimization. J Mach Learn Res 12:2121–2159

67. Dozat T (2016) Incorporating Nesterov momentum into Adam. In: International conference on learning representations (ICLR) workshop 2016, 2013–2016

68. Kingma DP, Ba J (2015) Adam: a method for stochastic optimization. In: Proceedings of the 3rd international conference for learning representations—ICLR (2015), San Diego, CA, USA, 7–9 May 2015

69. Kayci B, Demir BE, Demir F (2024) Deep learning based fault detection and diagnosis in photovoltaic system using thermal images acquired by UAV. J Polytech 27(1):91–99. https://doi.org/10.2339/politeknik.1094586

70. Huda ASN, Taib S (2014) A comparative study of MLP networks using backpropagation algorithms in electrical equipment thermography. Arab J Sci Eng 39:3873–3885. https://doi.org/10.1007/s13369-014-0989-7

71. Dhimish M, Badran G (2019) Photovoltaic hot-spots fault detection algorithm using fuzzy systems. IEEE Trans Device Mater Reliab 19(4):671–679. https://doi.org/10.1109/TDMR.2019.2944793

72. Al-Obaidy F, Yazdani F, Mohammadi FA (2017) Fault detection using thermal image based on soft computing methods: comparative study. Microelectron Reliab 71:56–64. https://doi.org/10.1016/j.microrel.2017.02.013

73. Zheng Q, Ma J, Liu M, Liu Y, Li Y, Shi G (2022) Lightweight hot-spot fault detection model of photovoltaic panels in UAV remote-sensing Image. Sensors 22(12):4617. https://doi.org/10.3390/s22124617

74. Dhimish M, Mather P, Holmes V (2019) Novel photovoltaic hot-spotting fault detection algorithm. IEEE Trans Device Mater Reliab 19(2):378–386. https://doi.org/10.1109/TDMR.2019.2910196

75. Venkatesh SN, Sugumaran V (2012) Fault Detection in aerial images of photovoltaic modules based on Deep learning. In: IOP conference series: materials science and engineering, Presented at the international conference on robotics, intelligent automation and control technologies (RIACT 2020) 2–3 Oct 2020, Chennai, India (2021)

76. Pierdicca R, Malinverni ES, Piccinini F, Paolanti M, Felicetti A, Zingaretti P (2018) Deep convolutional neural network for automatic detection of damaged photovoltaic cells. Int Arch Photogramm Remote Sens Spatial Inf Sci XLII–2:893–900

Publisher's Note