ISSN No:-2456-2165

Attention-Based Pre-Trained Model for Binary Image Classification on Small Datasets use Case -Glaucoma Image Classification

E Hukuimwe[1]; C Mafirabadza[2]; LNhapi[1]
Information Security and Assurance Department[1], Computer Science Department[2,3]
School of Information Science and Technology, Harare, Zimbabwe

Abstract:- Glaucoma is a prevalent eye disease that can lead to irreversible vision loss if not detected and treated early. Image classification techniques that make use of deep learning models have been showing promising results in diagnosing glaucoma.Traditional deep learning models often require large amounts of labeled data to achieve optimal performance.This paper explores the application of attention-based pretrained models for binary classification tasks using small datasets. However, in many real-world scenarios such as obtaining a substantial labeled dataset can be challenging or costly such as in rare diseses such as glaucoma.To address this issue, attention mechanisms have emerged as a powerful technique to enhance the performance of pretrained models by focusing on relevant features and samples. This paper investigates the effectiveness of attention-based pretrained models in the context of small datasets for binary classification tasks. Experimental results demonstrate that attention mechanisms can significantly improve the performance of pretrained models on limited data, making them a valuable tool for practical applications.

Keywords:- *Glaucoma, Deep Learning, Convolutional Neural Networks, Pretrained Models, Attention Mechanisms, Image Classification.*

I. INTRODUCTION

Before the evolvement of big data, smaller datasets such as CIFAR and NORB (Wei, 2019)which contains a few thousands of images were enough for machine learning models to learn basic recognition tasks. The large availability of data birthed to the rise of algorithms such as Alexnet and Googlnet which became an innovation on Convolution neural networks which were hard to apply on high resolution images (El-Dairi and House, 2019). A convolutional neural network (CNN), being the most recognized model for image recognition and classification was invented in the 1980s (J. C. Ye, 2022) and high-performance GPU-based CNN variants were used to achieve a recognition test error rate of 0.35%, 2.53% and 19.51% for digit recognition (MNIST), 3D object recognition (NORB), and natural images (CIFAR10) seemed to be the best performance(Cireşan *et al.*, 2011). In 2012(Krizhevsky, Sutskever and Hinton, 2017), Krizhevsky came with a model Alexnet which was trained on an ImageNet dataset which had 1.2 million high-resolution images.After fine tuning of the AlexNets hyper-parameters,ZFNet was created in 2013(Zeiler and Fergus, 2014) and ZFNet was said to have performed better than theAlexNet.GoogleNet came right after ZFNet in 2014 (Szegedy *et al.*, 2015), instead of having 8 layers as in AlexNet,GoogleNet had 22 CNN deep layers.ResNet was another vaiation of CNN which made use of the skip connection which fitted input from the previous layers without making any modifications on the input layer(He *et al.*, 2016) and its architecture had 152 CNN deep layers. The various CNN variations have confirmed the significance of depth that can lead to excellent performance of CNNs.

It is noted that the various variations of CNN models mentioned above were trained on the giant ImageNet datasets. Because of the rise of rare diseases such as glaucoma, small datasets are publicly accesseable. Efforts to collect large amount of data can be to no avail due to the low occurrences of such diseases and also privacy issues. There is need to come up with a technology that can efficiently work with small datasets inorder to achieve high performance

Attention based mechanism is introduced on a pretrained model in order to train a small dataset and giving emphasis to the most important information of an image .The proposed technique gives attention to relevant part of an image such that only specific parts of an image are used for example when trying to identify a dog in a picture that has grass and rocks, there is no need to give relevance to pixels that have grass and stones as they are not important in identifying a dog.The attention mechanism will be implemented on a pretrained model whereby a pretrained model is a saved network that is created by someone and trained on a large dataset in order to solve a similar problem. Instead of building a model from scratch, the use of a model trained on a similar problem can a be starting point and save the researcher cost and effort needed to gather ,clean and the infrastructure required to train the models. Thus, the use of an attention mechanism on the pretrained model to extract only relevant features needed for image classification for high performance.

II. BACKGROUND STUDY

The CNN models are a type of deep learning models that are used for image classification and recognition (Patil and Rane, 2021).The structure of the CNN model consist of input layers, hidden layers and then output layers. The input layer takes in raw data from the environment for example an image, the hidden layers takes the input data and apply convolutional filters to extract features from the image and

ISSN No:-2456-2165

finally the output layer that produces a prediction based on the features extracted .The development of CNN algorithms began with Yann Le Cuns backpropagation algorithm for handwriting recognition(LeCun *et al.*, 2012). Since then,there has been a lot of advances in the advancement of CNN models having Alexnet winning the ImageNet Large Scale Visual Recognition Challenge in 2012(Wei, 2019).This breakthrough by Alextnet demonstrated that deep learning can be used as state of art for computer vision tasks such as image classification and recognition. There has been various CNN model variations such as Googlenet, Resnet, Mobilenet after the birth of Alextnet.

('Techniques and pitfalls for ML training with small data sets - Trustbit - Accelerating Transformation', no date), stated that algorithms have become more effective as the datasets increases in size. Jana kube in his paper(Röglin *et al.*, 2022) argued that the use of small datasets has proved to be very difficult when dealing with rare diseases such as glaucoma due to the limited incidences therefore hard to sensitize clinical decision support systems to identify these diseases at an early stage. Efforts such as the use of data augmentation techniques has been used inorder to artificially increase the size of the dataset.However,the major limitation of data augmentation is data bias, i.e. the augmented data distribution can be quite different from the original one(Xu *et al.*, 2020). This data bias leads to a suboptimal performance of existing data augmentation methods and any deviations are not accepted in the medical field as they can result in the wrong classification and diagnosis.

It is noted in (Abdolrahimzadeh *et al.*, 2015) that glaucoma is a rare disorder that causes grave visual deterioration and it it usually occurs from birth to early ages of teenage. The paper went on to explain that 1 in 10,000 to 68,000 live births are diagnosed with glaucoma leading to very less incidents and therefore very small datasets. The largest dataset of glaucoma is the National Eye Institute's Glaucoma Database (NEI-GDB). This dataset contains over 1.5 million images from over 10,000 patients with glaucoma tht has glaucoma (Glaucoma Research Foundation, 2015). However not accessible for research leaving out publicly available datasets such as PAPILA dataset (Kovalyk *et al.*, 2022) with 448 images, G1020 (Li et al., no date) with 1835 images and RFUGE dataset [12] just to mention a few.

Having smaller datasets in glaucoma has made it difficult to apply the various variations of CNN models since they are data hungry. Various techniques on CNN models were used inorder to deal with unavailability of large datasets. Instead of using convolutional filters for feature extraction,image segmentation, Discreet Wavelet transform ,transfer learning and optimized deep learning models were used.The proposed model seeks to fine tune the convolution neural network and make use of an attention mechanism instead of convolutional filters.

III. RELATED WORK

lmage classification is a fundamental task in computer vision that involves assigning a label or category to an input image. Over the years, deep learning models have achieved remarkable success in image classification tasks, primarily due to their ability to learn hierarchical representations from large-scale datasets(Robert and Brown, 2004). However, these models often require massive amounts of labeled data for training, which may not always be available, especially in domains where data collection is expensive or time-consuming(Shanqing, Pednekar and Slater, 2019),(Shanqing, Pednekar and Slater, 2019).below are are some of the techniques used to deal with small datasets.In the effort to find a literature gap,the following methods are closely related to the proposed model.

A. Image Segmentation

Image segmentation,as mentioned in (Shu, 2019), was used in an CNN architecture as it does feature extaction by making use of the low level clustering features.Differently to (Shu, 2019), the proposed model model is pre-trained on a large scale datasets, using attention mechanism instead of feature extraction for image classification and finally implemting the modified model on a small dataset.

B. Transfer Learning

Transfer learning is a mechanism whereby one train data on huge dataset such as lmageNet and then extract features from the mechanism. It is meant to optimize performance, save time and the cost of recollecting, data. The paper (Pan and Yang, 2010),gave light on when to do transfer learning thus knowing the a situation that can work well with transfer learning, what to transfer that is getting acquinted with the part of information that needs to be transfere and finally one should get to know how to transfer depending on the how transfer issue. The main objective of transfer learning is to have high learning skills on the target task that is given. The reasercher went on to talk about negative transfer learning,a challenge that should be avoided among related tasks when doin transfer learning. Negative transfer exists if the relationship between source task and target task is weak. From this, we can note that it is importnat that bad information is recognized and rejected while learning a target task.Transfer learning is the improvement of learning in a new task based on the transfer ofknowledge from a related task that has already been learned (Gupta, Bhardwaj and Sharma, 2020). When training a new classifier with a small dataset there is high risk of overfitting and the new data will lead to poor generalization (Yao and Doretto, 2010).Dropout in(Srivastava *et al.*, 2014) and data augmentationas mentioned in(Xie *et al.*, 2015) , ('[PDF] Understanding Data Augmentation for Classification When to Warp Semantic Scholar', no date)are methods that are used in the effort to try to minimize overftting. The proposed model is similar to this method in the sense that there is the use of CNN architecture in the deep learning models such as Alexnet on image classification based on large scale datasets, (Chan *et al.*, 2015), and(Cires *et al.*, 2013).Similar to the above methods of classification, Mengying Shu (RI, 2019) tried to pretrain the model to fit small datasets by fine tuning the parameters of the existing

model inorder to fit small datsets .Differently from the proposed model , pretraining is done thus making use of transfer learning but use attention mechanism on an CNN model for image classification unlike using the feature extraction used on the CNN architecture of the deep learning algortihms used.

C. *Deep Learning*

Deep learning allows models that are composed of multiple layers which are used to learn data (Solomon *et al.*, 2004). Examples of deep learning algorithms are AlexNet (Shu, 2019) , VGG net (Fergus and Road, no date), GoogleNet (Szegedy *et al.*, 2015), ResNet (Ren *et al.*, 2017) and YOLO (You Only Look Once) . Considering deep learning algorithms, Alexnet is going to be used. While it is shown that all of the models work well in practice, it is unclear that those models perform well when modified and used to fit the small datasets. A deep learning CNN algorithm Alexnet, was modified in order to do image classification on a Tiny Imag, a subset of the Imagenet dataset(Lucas, 2020). The neurons of the layers were reduced in order to fit a small dataset thus producing a reduced fine tuned Alexnet .The modified alexnet was said to have performed well but experienced overfitting as the number of epochs increased. The idea was to prove that these algorithms can work on small datasets when proper modifications are done. Similar to the proposed model going pretraining is going to be used . Differently from the proposed model, project , attention mechanism is going to be used instead of feature extraction used in the CNN architecture of the alexnet.

In 2021,Axel Masquelin worked on a image classification model using a modified CNN architecture for classification of medical image (Masquelin *et al.*, 2022).In his paper he alluded that due to the lack of large medical datasets ,there is high overfiiting and generalization .Axel proposed the use of Discreet Wavelemghth transform an image processing technique that samples wavelets at discret intervals.Before Discreet Wavelet was implemented ,feature extraction was used to efficiently perfom edge scale normalization.When Discreet wavelength was used instead of convolutional operations, the DWT perfomed better with AUC of 94% and 92 % respectively Similiar to the proposed model, CNN architecture is going to be used but using attention mechanism for feature extraction while this model used Discreet Wavelet Trasform with edge scale normalization.

Overally ,the above techniques that were implemted inorder to deal with small datasets used Feature extraction on the CNN algorithms which was stated in a 2019 onDiscreet wavelet transform paper that CNN models which uses feature extraction and selection are time consuming and they varies depending on different types of objects thus the need to come up with a CNN using attention mechanism.

D. *Attention mechanisms*

Attention mechanisms are a fundamental component of many machine learning models, particularly in the field of natural language processing (NLP)(Wu *et al.*, 2018). They enable models to focus on specific parts of the input data that are most relevant for making predictions or generating output.The concept of attention is inspired by human cognitive processes, where attention allows us to selectively process and prioritize certain information while ignoring others. Similarly, attention mechanisms in machine learning models help allocate resources and computational power to relevant parts of the input(Sanocki and Lee, 2022).

The core idea behind attention mechanisms is to compute a set of attention weights that indicate how much each element in the input should contribute to the model's decision-making process. These weights are typically calculated by comparing each element's relevance with respect to a specific query or context vector(Choi and Lee, 2023),.

There are different types of attention mechanisms, such as additive attention and multiplicative attention. Additive attention computes the relevance between query and key vectors by applying a feed-forward neural network layer(Vaswani, 2017). Multiplicative attention, on the other hand, calculates relevance through dot product or cosine similarity between query and key vectors.

Once the attention weights are computed, they are used to weight the values associated with each element in the input sequence. These weighted values are then combined (usually summed) to produce an aggregated representation called the context vector. The context vector is then used by the model for making predictions or generating output(Vaswani, 2017).

Attention mechanisms have significantly improved various NLP tasks by allowing models to focus on relevant information while ignoring noise or irrelevant details(Wu *et al.*, 2018). They have also enabled better interpretability and explainability of model predictions since it is possible to visualize which parts of the input were attended more heavily during inference.

Overall, attention mechanisms have become a crucial tool in modern machine learning models, enabling them to handle complex and context-dependent tasks more effectively.

IV. METHODOLOGY

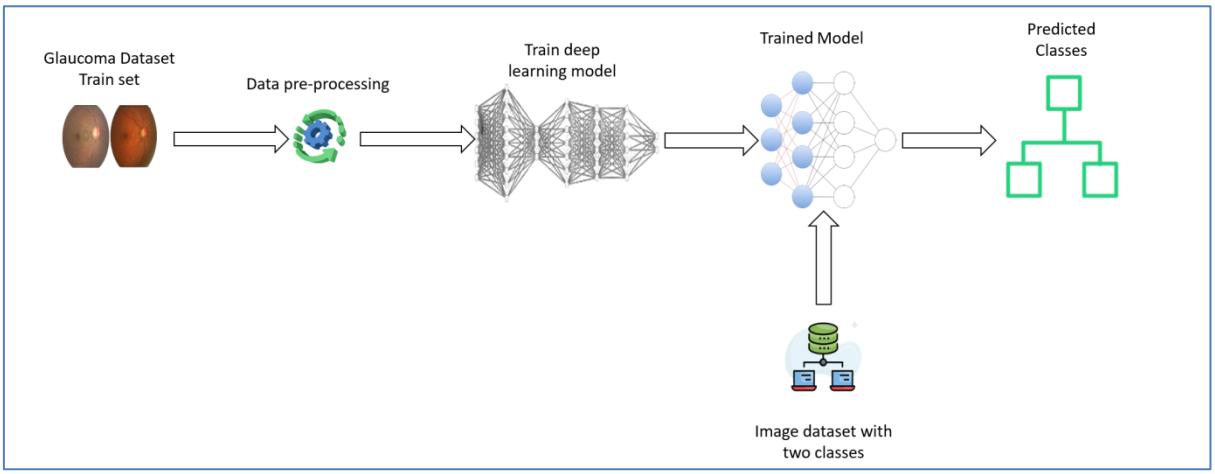

Fig. 1: Classification process flow

In this section we introduce the method used in the design and implementation of a pre-trained attention based model for binary image classification. A deep learning CNN based model is trained using Glaucoma dataset and fine tuned on the dataset to achieve higher accuracy. Multi-head Attention (Wang , 2023) is added to the model to reduce number of input features to the model and concentrate more on the features that distinguish negative and positive cases for glaucoma.

The pre-trained model is then used for classfyiing other datasets in the medical field which have binary classes and they have small datasets available.

A. Dataset Selection:

In this section, dataset collection and preprocessing steps are used for fine-tuning the fundus image for further processing. The database contains 4,584 fundus images with 1,711 positive and 3,143 negative glaucoma samples obtained from Beijing Tongren Hospital. Each fundus image is diagnosed by qualified glaucoma specialists, taking the consideration of both morphologic and functional analysis, iglaucoma affects mainly the optic disc and the area near it. Therefore, most prior works suggest that only the optic disc part of the image should be used. However, we found that this part of the image can be very small in many images and convey very little signal when the image contrast is not set correctly. We explored different alternatives, to establish what part of the image is better suited to detect glaucoma and found that using the full image reached best results. The only preprocessing performed consists in removing the black borders from the input image.

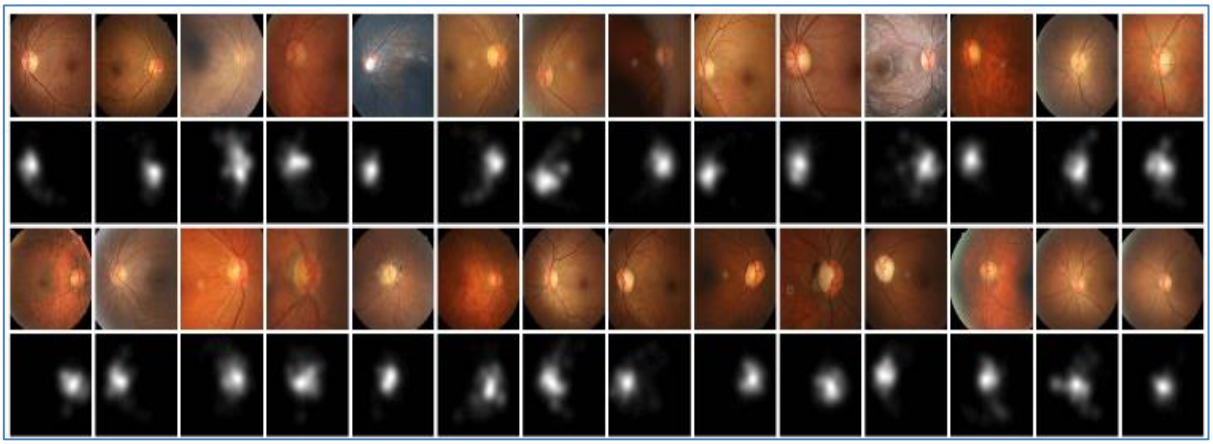

Fig. 2: Fundus images of abnormal glaucoma samples

The preprocessing technique plays a crucial role in image processing by ensuring that all images are standardized before undergoing the actual analysis. Its purpose is to enhance the quality of the data images by eliminating unwanted elements such as speckles, blind spots, noise, low contrast, and irrelevant variations. This preprocessing scheme aims to improve essential aspects for subsequent processing. It involves various procedures such as image resizing, channel extraction, noise removal, and image enhancement, as demonstrated in Figure 3.

ISSN No:-2456-2165

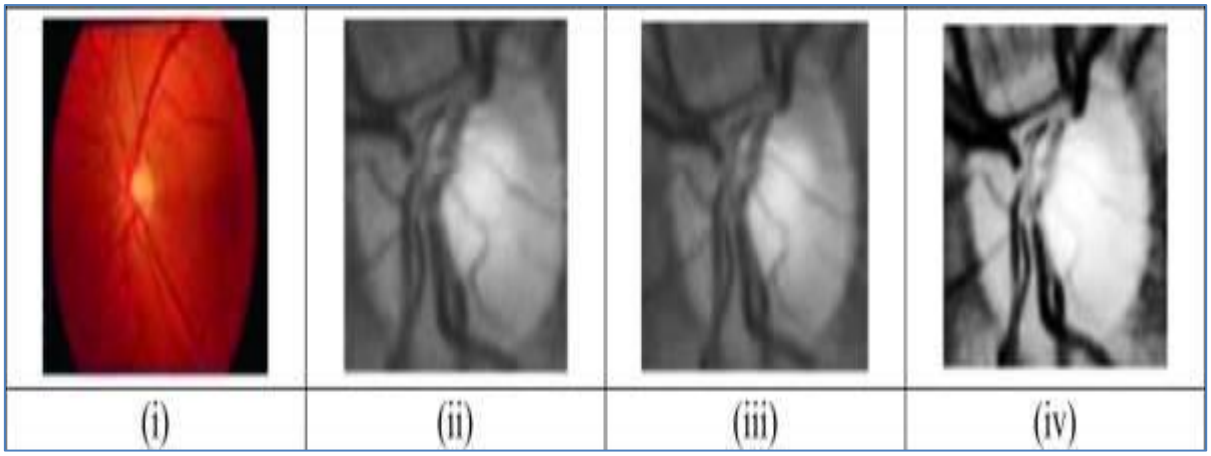

Fig. 3: Preprocessing output

Preprocessing output: (i) input image, (ii) green channel-extracted image, (iii) median-filtered image, and (iv) CLAHE image

To facilitate better analysis, the input fundus and OCT images obtained from the database are resized to a uniform resolution of 300 × 300 pixels. This resizing process ensures that all original images have the same dimensions, resolution, and scale, thereby enhancing comparability and facilitating more effective analysis.

B. Model Selection and Training

The goal of the image classification stage is to categorize an input image into one of two categories: glaucoma positive and glaucoma negative, using a deep learning CNN with a multi-head attention algorithm. To achieve this, the image classification process is divided into three sequential steps: CNN model selection, experimental evaluation, and ensemble construction, and the analysis of results.

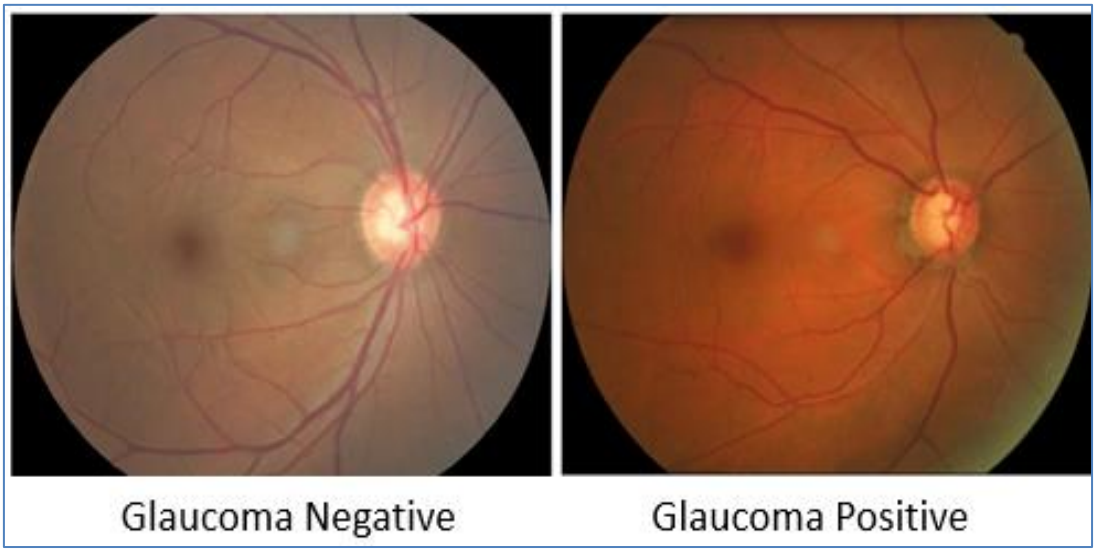

Fig. 4: Glaucoma images examples of Positive and Negative

For image classification tasks, an attention-based deep learning model architecture can be utilized. It is recommended to split the cleaned dataset into training and validation sets. The training set is used to train the deep learning model, optimizing its parameters to minimize the loss function. The validation set is used to evaluate the model's performance. If necessary, the model can be fine-tuned based on the results obtained during validation.

Fig 5 shows the architecture of the model

ISSN No:-2456-2165

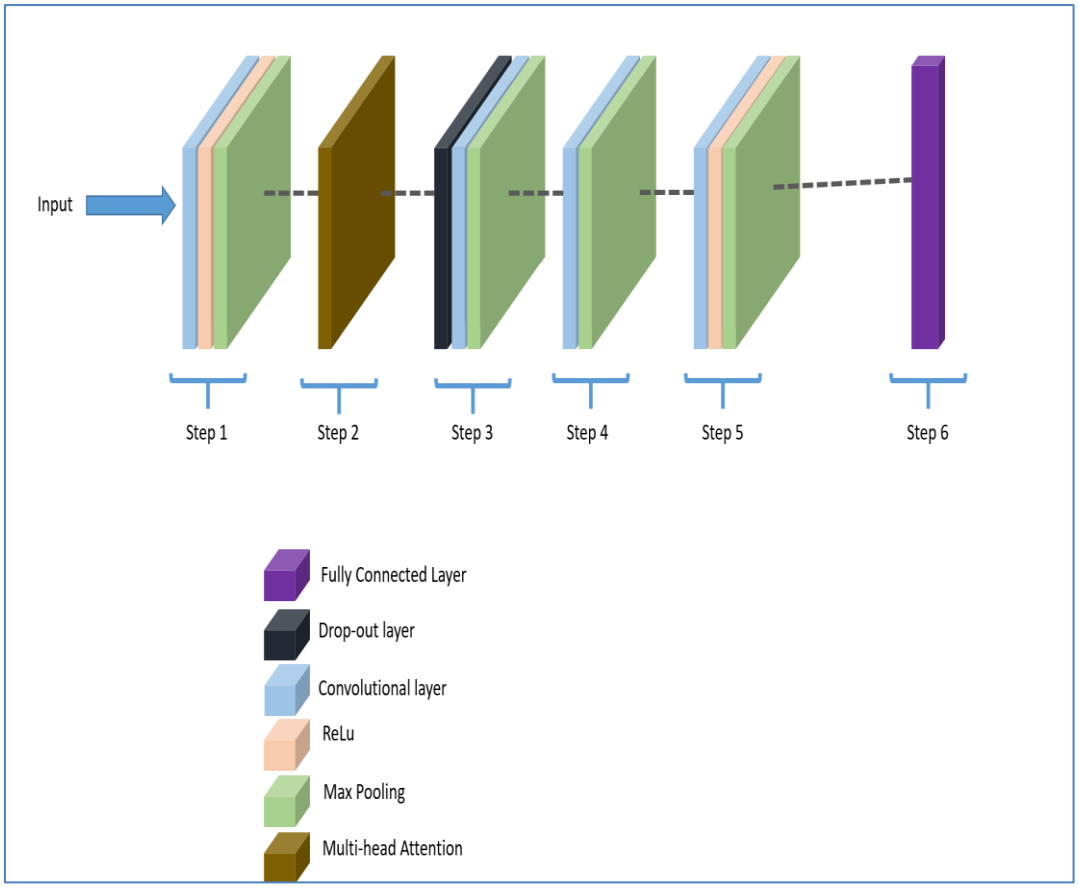

Fig. 5: model architecture

Step 1: The are three layers, convolutional , ReLU, and Max pooling layer. The convolutional layers consist of 64 3X3 filters with a stride of 1 and a padding of 1. The ReLU acts as the activation function to remove linearity in the function.

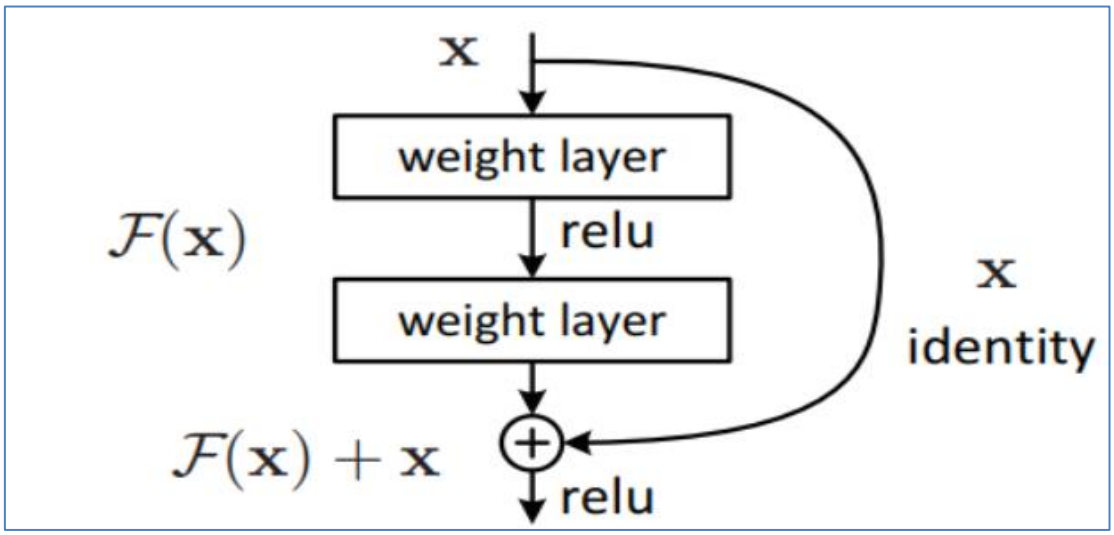

Fig. 6: ReLU activation function

Step 2: Multi-head Attention:

ISSN No:-2456-2165

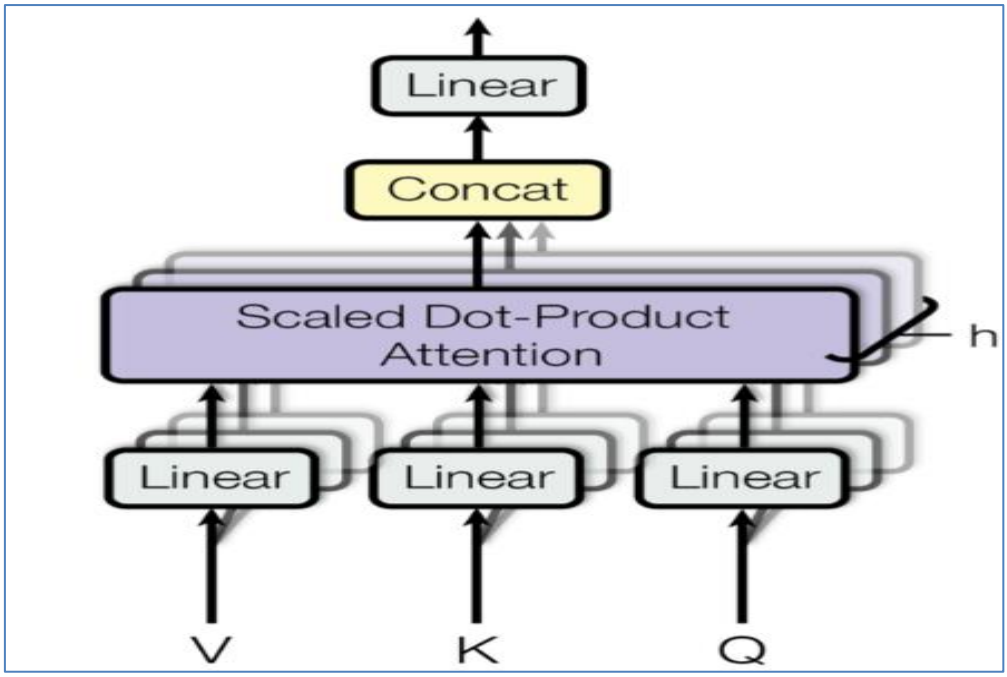

Fig. 7: Multi-head attention mechanism

Multi-head Attention is a component used in attention mechanisms, where it performs multiple iterations of an attention mechanism simultaneously. The individual outputs from each independent attention iteration are then combined by concatenating them and linearly transforming them to achieve the desired output dimension. The concept behind using multiple attention heads is to enable differential attention to different parts of the sequence, allowing for more nuanced and focused analysis of the input.

C. Evaluation Metrics

The following metrics were used in this model:

- **Accuracy:** expressing the number of correct predictions that are done out of the total predictions expressed as a percentage

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

- **Precision:** the precision is goin to be measured by taking samples belonging to the same class that were correctly classified and are in comparison with the samples tha were predicted positively classified.

$$\text{Precision} = \frac{\sum TP}{\sum TP + \sum FN}$$

- **Recall:** Recall or true positive rate (tpr) is an indicator of a classifier's capacity to correctly pick instances of the target class related to the positive samples.

$$\text{Recall} = \frac{\sum TP}{\sum TP + \sum FP}$$

- **F-measure:** The f-measure is the harmonic mean of precision and recall.

$$\text{F1 Score} = \frac{2 \times Pr \times Re}{Pr + Re}$$

In the above equations, TP, TN, FP, and FN represent the true positive, true negative, false positive, and false negative, respectively. Pr represents precision and Re represents Recall.

V. EXPERIMENTAL RESULTS

This section presents the experimental results that aim to validate the effectiveness of our approach in detecting glaucoma and localizing pathological areas. For the experiment, we utilized a total of 4,584 fundus images from our LAG database. These images were randomly divided into training (3,584 images) and validation (1000 images) sets.

In this research, deep learning algorithm was employed using convolutional neural network (CNN) model that was pre-trained on the LAG dataset. This utilization of pre-training enabled transfer learning, leveraging the knowledge gained from the LAG dataset to improve performance in the current study.

To adapt the CNN classifiers to the new dataset, we made necessary configurations by adjusting their parameters prior to training. This involved a process known as weight freezing, where we preserved the weights and knowledge acquired during pre-training on the ImageNet dataset. We froze a portion of the model while introducing two new trainable layers on top of the frozen layers. Subsequently, we trained these new layers using the training images from the LAG dataset as input. Figure 8 provides a visual representation of this process.

ISSN No:-2456-2165

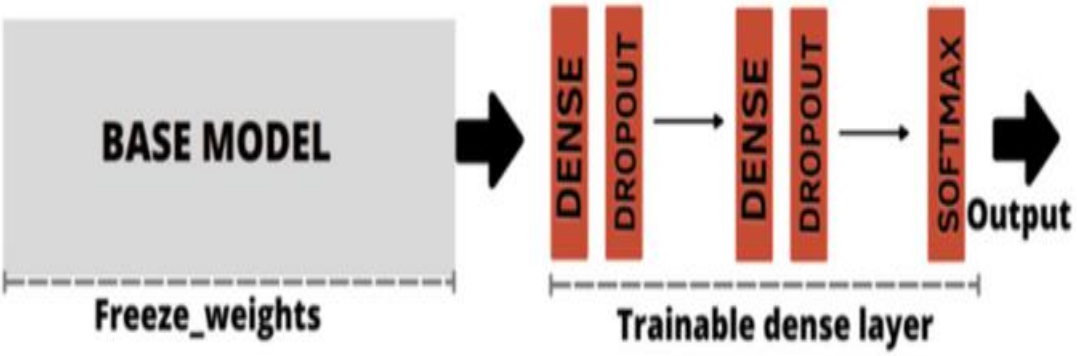

Fig. 8: Trainable dense layer

Table 1 displays the four CNN models that were specifically chosen for this research. These classifiers were selected to compare the effect of inclusing attention and the effect of using transfer learning.

Table 1: Results Comparison

Model	Accuracy	Sensitivity	Specificity
CNN	98.74	72.54	98.47
CNN with Attention	98.88	70.83	98.78
Pre-trained CNN	98.9	73.02	98.8
Pre-trained CNN with Attention	99.12	84.62	99.16

The proposed approach for the pre-training and attention gives 99.12% accuracy, 84.62% sensitivity, and 99.16% specificity. Table Table 1 depicts the comparison performances of the proposed method and three other methods used. The models were all first trained on the LAG dataset , the datset was also then fed as new datset for classification. Combining pre-training and attention outperformed all the other methods.

VI. CONTRIBUTIONS

When developing this attention-based model for binary image classification in small datasets, several contributions were made to enhance its effectiveness and address the challenges associated with limited data. Here are the contributions:

- Attention Mechanisms:Introduce attention mechanisms into the model architecture to enable the model to focus on informative regions within the images. Attention mechanisms allow the model to assign different weights to different parts of the image, emphasizing the most relevant features for accurate classification.
- Regularization Techniques: Incorporate regularization techniques to prevent overfitting and improve generalization. Techniques such as dropout, batch normalization, or L1/L2 regularization are utilized to reduce the model's sensitivity to variations in the training dataset and improve its ability to generalize to unseen data.
- Model Interpretability:Develop methods to interpret and visualize the attention weights assigned by the model. This can provide insights into the regions of the image that are crucial for classification decisions. Techniques such as saliency maps or Grad-CAM can be used to visualize the attention regions and aid in understanding the model's decision-making process.
- Cross-Validation and Hyperparameter Tuning: Perform rigorous cross-validation to evaluate the model's performance and assess its generalization capabilities. Hyperparameter tuning techniques, such as grid search or Bayesian optimization, can be applied to optimize the model's parameters and find the best configuration for the specific dataset and task.

VII. CONCLUSION

In this paper, we present an introduction to a pre-trained attention-based deep learning model designed for the binary classification of rare diseases. This model addresses the challenge of limited datasets typically encountered in cases such as rare dieases like glaucoma. Specifically, our model was trained on the LAG dataset, which consists of 4,584 glaucoma images. To emphasize the distinguishing features between glaucoma-positive and glaucoma-negative cases, we employed a multi-head Attention mechanism.The model comprises subnets for attention prediction, enabling the detection of glaucoma by highlighting deep features through visualized maps of pathological areas based on predicted attention maps. To assess the effectiveness of our method, we conducted several experiments, (1)Utilizing a CNN model to classify glaucoma images from the LAG dataset, (2)Utilizing a CNN with attention to classify glaucoma images from the LAG dataset, (3)Employing a pre-trained CNN to classify glaucoma images from the LAG dataset,(4)Utilizing an attention-based pre-trained model to classify glaucoma images from the LAG dataset.The experimental results demonstrate that our pre-trained attention-based model outperformed the other three methods. In future work, we intend to evaluate our model on new datasets to assess its generalizability across different datasets featuring binary classes for diseases.

REFERENCES

[1]. '[PDF] Understanding Data Augmentation for Classification When to Warp Semantic Scholar' (no date).

[2]. '3. GoogleNet' (2014), p. 22.

[3]. Abdolrahimzadeh, S. *et al.* (2015) 'Rare Diseases Leading to Childhood Glaucoma: Epidemiology, Pathophysiogenesis, and Management', *BioMed Research International*, 2015. Available at: https://doi.org/10.1155/2015/781294.

[4]. Alom, M.Z. *et al.* (2019) 'A state-of-the-art survey on deep learning theory and architectures', *Electronics (Switzerland)*, 8(3), pp. 1–67. Available at: https://doi.org/10.3390/electronics8030292.

[5]. Chan, T.H. *et al.* (2015) 'PCANet: A Simple Deep Learning Baseline for Image Classification?', *IEEE Transactions on Image Processing*, pp. 5017–5032. Available at: https://doi.org/10.1109/TIP.2015.2475625.

[6]. Choi, S.R. and Lee, M. (2023) 'Transformer Architecture and Attention Mechanisms in Genome Data Analysis : A Comprehensive Review'.

[7]. Cires, D.C. *et al.* (2013) 'Flexible, High Performance Convolutional Neural Networks for Image Classification', *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence Flexible*, pp. 1237–1242. Available at: https://www.aaai.org/ocs/index.php/IJCAI/IJCAI11/paper/viewFile/3098/3425.

[8]. Cireşan, D.C. *et al.* (2011) 'High-Performance Neural Networks for Visual Object Classification'. Available at: http://arxiv.org/abs/1102.0183.

[9]. Datta, S.K. and Gao, M. (2021) 'Soft-Attention Improves Skin Cancer Classification Performance', pp. 1–8.

[10]. Du, K.L. and Swamy, M.N.S. (2014) *Neural networks and statistical learning, Neural Networks and Statistical Learning*. Available at: https://doi.org/10.1007/978-1-4471-5571-3.

[11]. El-Dairi, M. and House, R.J. (2019) 'Optic nerve hypoplasia', *Handbook of Pediatric Retinal OCT and the Eye-Brain Connection*, pp. 285–287. Available at: https://doi.org/10.1016/B978-0-323-60984-5.00062-7.

[12]. Fergus, R. and Road, P. (no date) 'Fergus03.Pdf'.

[13]. Fezari, M., Dahoud, A. Al and Al-dahoud, A. (2023) 'State of the Art of Deep Neural Networks Models', (May).

[14]. Ghosh, A. *et al.* (2019) *Fundamental concepts of convolutional neural network, Intelligent Systems Reference Library*. Available at: https://doi.org/10.1007/978-3-030-32644-9_36.

[15]. Glaucoma Research Foundation (2015) 'Glaucoma Facts and Stats', *Glaucoma Research Foundtion*, pp. 11–13. Available at: https://www.glaucoma.org/glaucoma/glaucoma-facts-and-stats.php.

[16]. Gupta, R., Bhardwaj, K.K. and Sharma, D.K. (2020) 'Transfer Learning', *Machine Learning and Big Data: Concepts, Algorithms, Tools and Applications*, pp. 337–360. Available at: https://doi.org/10.1002/9781119654834.ch13.

[17]. Hanif, M.S. and Bilal, M. (2020) 'Competitive residual neural network for image classification', *ICT Express*, 6(1), pp. 28–37. Available at: https://doi.org/10.1016/j.icte.2019.06.001.

[18]. He, K. *et al.* (2016) 'Deep residual learning for image recognition', *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-Decem, pp. 770–778. Available at: https://doi.org/10.1109/CVPR.2016.90.

[19]. Jin, Y. *et al.* (2020) 'Multi-Head Self-Attention-Based Deep Clustering for Single-Channel Speech Separation', *IEEE Access*, 8, pp. 100013–100021. Available at: https://doi.org/10.1109/ACCESS.2020.2997871.

[20]. Kovalyk, O. *et al.* (2022) 'PAPILA: Dataset with fundus images and clinical data of both eyes of the same patient for glaucoma assessment', *Scientific Data*, 9(1), pp. 1–13. Available at: https://doi.org/10.1038/s41597-022-01388-1.

[21]. Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2017) 'ImageNet classification with deep convolutional neural networks', *Communications of the ACM*, 60(6), pp. 84–90. Available at: https://doi.org/10.1145/3065386.

[22]. Kulkarni, M. *et al.* (2020) 'Soft Attention Convolutional Neural Networks for Rare Event Detection in Soft Attention Convolutional Neural Networks for Rare Event Detection in Sequences', (November).

[23]. LeCun, Y.A. *et al.* (2012) 'Efficient BackProp BT - Neural Networks: Tricks of the Trade', *Neural Networks: Tricks of the Trade*, 7700(Chapter 3), pp. 9–48. Available at: http://link.springer.com/chapter/10.1007/978-3-642-35289-8_3/fulltext.html%5Cnpapers3://publication/doi/10.1007/978-3-642-35289-8_3.

[24]. Li et al. (no date) 'CUHK03 Dataset | Papers With Code'. Available at: https://paperswithcode.com/dataset/cuhk03.

[25]. Li, Z. *et al.* (2021) 'Local attention sequence model for video object detection', *Applied Sciences (Switzerland)*, 11(10), pp. 1–10. Available at: https://doi.org/10.3390/app11104561.

[26]. Lindsay, G.W. (2020) 'Attention in Psychology, Neuroscience, and Machine Learning', *Frontiers in Computational Neuroscience*, 14(April), pp. 1–21. Available at: https://doi.org/10.3389/fncom.2020.00029.

[27]. Lucas, E. (2020) 'Optimizing AlexNet on Tiny ImageNet10', pp. 1–10.

[28]. Mahdizadehaghdam, S., Panahi, A. and Krim, H. (2019) 'Sparse generative adversarial network', *Proceedings - 2019 International Conference on Computer Vision Workshop, ICCVW 2019*, pp. 3063–3071. Available at: https://doi.org/10.1109/ICCVW.2019.00369.

[29]. Marhon, S.A., Cameron, C.J.F. and Kremer, S.C. (2013) 'Recurrent Neural Networks', *Intelligent Systems Reference Library*, 49, pp. 29–65. Available at: https://doi.org/10.1007/978-3-642-36657-4_2.

[30]. Masquelin, A.H. *et al.* (2022) 'for Medical Image Classification by Deep Learning', 155(2), pp. 309–317. Available at: https://doi.org/10.1007/s00418-020-01961-y.Wavelet.
[31]. Nadim, U.S.M. and Jung, Y.J. (2020) 'Global and local attention-based free-form image inpainting', *Sensors (Switzerland)*, 20(11), pp. 1–27. Available at: https://doi.org/10.3390/s20113204.
[32]. Pan, S.J. and Yang, Q. (2010) 'A survey on transfer learning', *IEEE Transactions on Knowledge and Data Engineering*, 22(10), pp. 1345–1359. Available at: https://doi.org/10.1109/TKDE.2009.191.
[33]. Patil, A. and Rane, M. (2021) 'Convolutional Neural Networks: An Overview and Its Applications in Pattern Recognition', *Smart Innovation, Systems and Technologies*, 195, pp. 21–30. Available at: https://doi.org/10.1007/978-981-15-7078-0_3.
[34]. 'Proposed-residual-network-architecture-The-output-dimensions-at-each-single-layer-are' (no date).
[35]. Ren, S. *et al.* (2017) 'Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), pp. 1137–1149. Available at: https://doi.org/10.1109/TPAMI.2016.2577031.
[36]. RI, M.K. (2019) 'No TitleΕΛΕΝΗ', *Αγαη*, 8(5), p. 55.
[37]. Robert, B. and Brown, E.B. (2004) 'No 主観的健康感を中心とした在宅高齢者における健康関連指標に関する共分散構造分析Title', (1), pp. 1–14.
[38]. Röglin, J. *et al.* (2022) 'Improving classification results on a small medical dataset using a GAN; An outlook for dealing with rare disease datasets', *Frontiers in Computer Science*, 4. Available at: https://doi.org/10.3389/fcomp.2022.858874.
[39]. Salavati, P. and Mohammadi, H.M. (2018) 'Obstacle detection using GoogleNet', *2018 8th International Conference on Computer and Knowledge Engineering, ICCKE 2018*, (October 2018), pp. 326–332. Available at: https://doi.org/10.1109/ICCKE.2018.8566315.
[40]. Sangeetha, V. and Prasad, K.J.R. (2006) 'Syntheses of novel derivatives of 2-acetylfuro[2,3-a]carbazoles, benzo[1,2-b]-1,4-thiazepino[2,3-a]carbazoles and 1-acetyloxycarbazole-2- carbaldehydes', *Indian Journal of Chemistry - Section B Organic and Medicinal Chemistry*, 45(8), pp. 1951–1954. Available at: https://doi.org/10.1002/chin.200650130.
[41]. Sanocki, T. and Lee, J.H. (2022) 'Attention-Setting and Human Mental Function'.
[42]. 'Schematic-diagram-of-a-basic-convolutional-neural-network-CNN-architecture-26' (no date).
[43]. Shamir, A. (no date) 'No Title'.
[44]. Shanqing, G., Pednekar, M. and Slater, R. (2019) 'Improve Image Classification Using Data Augmentation and Neural Networks', *SMU Data Science Review*, 2(2), pp. 1–43. Available at: https://scholar.smu.edu/datasciencereviewhttp://digitalrepository.smu.edu.Availableat:https://scholar.smu.edu/datasciencereview/vol2/iss2/1.
[45]. Shu, M. (2019) 'Deep learning for image classification on very small datasets using transfer learning', *Creative Components*, pp. 14–21.
[46]. Solomon, T. *et al.* (2004) 'No 主観的健康感を中心とした在宅高齢者における健康関連指標に関する共分散構造分析Title', *International Journal of Tropical Insect Science*, 8(4), pp. 104–110.
[47]. Srivastava, N. *et al.* (2014) 'Dropout: A simple way to prevent neural networks from overfitting', *Journal of Machine Learning Research*, 15, pp. 1929–1958.
[48]. Szegedy, C. *et al.* (2015) 'Going deeper with convolutions', *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 07-12-June, pp. 1–9. Available at: https://doi.org/10.1109/CVPR.2015.7298594.
[49]. Tan, Y.F. *et al.* (2023) 'Human activity recognition with self-attention', *International Journal of Electrical and Computer Engineering*, 13(2), pp. 2023–2029. Available at: https://doi.org/10.11591/ijece.v13i2.pp2023-2029.
[50]. 'Techniques and pitfalls for ML training with small data sets - Trustbit - Accelerating Transformation' (no date).
[51]. Teuwen, J. and Moriakov, N. (2019) *Convolutional neural networks*, *Handbook of Medical Image Computing and Computer Assisted Intervention*. Elsevier Inc. Available at: https://doi.org/10.1016/B978-0-12-816176-0.00025-9.
[52]. Udrea, M. and Strisciuglio, N. (no date) 'A Comparative Study on Pre-trained Classifiers in the Context of Image Classification', pp. 1–8.
[53]. Vaswani, A. (2017) 'Attention Is All You Need', (Nips).
[54]. Vision, C. *et al.* (2020) 'What is Resnet or Residual Network | How Resnet Helps? Introduction to Resnet or Residual Network', pp. 1–8. Available at: https://www.mygreatlearning.com/blog/resnet/.
[55]. Wang Tingwu (no date) 'Contents 1. Why do we need Recurrent Neural Network?', p. 41. Available at: https://www.cs.toronto.edu/~tingwuwang/rnn_tutorial.pdf.
[56]. Wei, J. (2019) 'AlexNet: The Architecture that Challenged CNNs | by Jerry Wei | Towards Data Science', *Towards Data Science* [Preprint]. Available at: https://towardsdatascience.com/alexnet-the-architecture-that-challenged-cnns-e406d5297951.
[57]. Wu, L. *et al.* (2018) 'Word attention for sequence to sequence text understanding', *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*, pp. 5578–5585. Available at: https://doi.org/10.1609/aaai.v32i1.11971.
[58]. Xie, S. *et al.* (2015) 'Hyper-class augmented and regularized deep learning for fine-grained image classification', *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 07-12-June, pp. 2645–2654. Available at: https://doi.org/10.1109/CVPR.2015.7298880.
[59]. Xu, Y. *et al.* (2020) 'WeMix: How to Better Utilize Data Augmentation'.
[60]. Yao, Y. and Doretto, G. (2010) 'Boosting for transfer learning with multiple sources', *Proceedings of the*

IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1855–1862. Available at: https://doi.org/10.1109/CVPR.2010.5539857.

[61]. Ye, J.C. (2022) 'Convolutional Neural Networks', *Mathematics in Industry*, 37(Icectt), pp. 113–134. Available at: https://doi.org/10.1007/978-981-16-6046-7_7.

[62]. Ye, Y. (2022) 'Generative adversarial networks', p. 28. Available at: https://doi.org/10.1117/12.2626949.

[63]. Zeiler, M.D. and Fergus, R. (2014) 'Visualizing and understanding convolutional networks', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8689 LNCS(PART 1), pp. 818–833. Available at: https://doi.org/10.1007/978-3-319-10590-1_53.

[64]. Zeng, G. *et al.* (2016) 'Preparation of novel high copper ions removal membranes by embedding organosilane-functionalized multi-walled carbon nanotube', *Journal of Chemical Technology and Biotechnology*, 91(8), pp. 2322–2330. Available at: https://doi.org/10.1002/jctb.4820.

[65]. Zhang, Z. *et al.* (2020) 'Relation-Aware Global Attention for Person Re-Identification', *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 3183–3192. Available at: https://doi.org/10.1109/CVPR42600.2020.00325.

[66]. Wang, H., Xu, J., Yan, R., Sun, C. and Chen, X., 2020. Intelligent bearing fault diagnosis using multi-head attention-based CNN. *Procedia Manufacturing*, *49*, pp.112-118.